TOWARDS QUANTIFYING UPPER-ARM REHABILITATION METRICS FOR CHILDREN THROUGH INTERACTION WITH A HUMANOID ROBOT

A Thesis
Presented to
The Academic Faculty

by

Douglas A. Brooks

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Electrical and Computer Engineering

Georgia Institute of Technology
August 2012

TOWARDS QUANTIFYING UPPER-ARM REHABILITATION METRICS FOR CHILDREN THROUGH INTERACTION WITH A HUMANOID ROBOT

Approved by:

Professor Patricio Vela,
Committee Chair
School of Electrical and Computer Engineering
Georgia Institute of Technology

Professor Pamela Bhatti
School of Electrical and Computer Engineering
Georgia Institute of Technology

Professor Ayanna M. Howard, Advisor
School of Electrical and Computer Engineering
Georgia Institute of Technology

Professor Charlie Kemp
Department of Biomedical Engineering
Georgia Institute of Technology

Professor Linda Wills
School of Electrical and Computer Engineering
Georgia Institute of Technology

Date Approved: April 6, 2012

First, I would like to thank my Lord and Savior Jesus Christ. No part of this journey would have been possible without You. Second, I would like to thank each member of family for his or her continued support: Ethelene Mc Leod, John Mc Leod, Tonnette Mc Leod, Thomas Mc Leod, Dwayne Mc Leod, Doug Brooks, Crystal Lewis, and everyone else. I am also grateful to have had a wonderful mentor and friend in Dr. Brad Jones for gudiance during this process. I am also deeply in debt to the love and support given by Dr. Shara McClendon. For Simone and Zion Brooks, dream big, reach high, but most of all, trust God. Finally, this work is in loving memory of J.R. Price, Harold Mills jr., Aaron McLean, and Jimmy Brooks.

ACKNOWLEDGEMENTS

This work was supported in part by the National Science Foundation under Grant CNS-0958487. Any opinion, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. I would also like to express my gratitude to my collaborators Dr. Edward Brown, Rochester Institute of Technology, for providing assistance in contact rehabilitation therapy, and Dr. Yu-ping Chen, Georgia State University, for providing assistance with IRB approvals, the use of the Vicon Motion Capture System, and continued aid with regards to physical therapeutic exercises and terminology. I would also like to acknowledge each member of the Human Automation Systems Lab for his or her continued insight regarding the approaches presented here as well as continued support.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS OR ABBREVIATIONS

AROM Active Range of Motion.

CP Cerebral Palsy.

EMG Electromyography.

FM Fine Motor.

GM Gross Motor.

GMFM Gross Motor Function Measure.

GP Gaussian Process.

HMM Hidden Markov Model.

IR Infrared.

MHI Motion History Image.

MM Multimodal Mean.

MT Movement Time.

MU Movement Units.

NJS Normalized Jerk Score.

PAV Peak Angular Velocity.

PCA Principle Component Analysis.

PDMS Peabody Developmental Motor Scales.

PT Physical Therapist.

PV Peak Veolcity.

RGB Red Green Blue.

Stdv Standard Deviation.

TD Total Displacement.

VMCS Vicon Motion Capture System.

SUMMARY

The objective of this research effort is to further rehabilitation techniques for children by developing and validating the core technologies needed to integrate therapy instruction with child-robot play interaction in order to improve upper-arm rehabilitation. Using computer vision techniques such as Motion History Imaging (MHI), Multimodal Mean, edge detection, and Random Sample Consensus (RANSAC), movements can be quantified through robot observation. Also incorporating three-dimensional data obtained via an infrared projector coupled with a Principle Component Analysis (PCA), depth information can be utilized to create a robust algorithm. Finally, utilizing prior knowledge regarding exercise data, physical therapeutic metrics, and novel approaches, a mapping to therapist instructions can be created allowing robotic feedback and intelligent interaction.

CHAPTER I

INTRODUCTION

Rehabilitation therapy can be a very practitioner intensive process. When patients enter into the process they are often required or asked to perform exercises that they have been shown how to do when they are at home between visits. Proper compliance is strongly correlated with shorter time to recovery as well as reduction of pain in the long term [111]. During the time between therapy sessions there are many factors which affect patient compliance, including forgetfulness, lack of motivation, boredom, and lack of instant feedback. To deal with these issues, researchers have shown the positive use of robots in assistive therapy applications ranging from stroke rehabilitation [15, 47, 49, 60, 63, 64, 81, 90, 109] to motor development in children [46]. In this venue, robotic systems for rehabilitation have generally been used to record information about the motor performance (position, trajectory, interaction force/impedance) during active movements [30]. These systems have been used to objectively assess the performance of a patient through repeatable and quantifiable metrics, as an effective means for rehabilitation therapy [10, 21, 40, 50]. The major barrier is that, to date, most assistive robotic devices are not designed for children, both neurotypical and non-neurotypical. This causes a unique challenge for deploying such robotics for this target demographic.

To overcome this barrier, state-of-the-art techniques must be created to facilitate the interaction necessary to be useful for therapeutic rehabilitation with respect to children. Although state-of-the-art in robotic systems for children with disabilities have piloted their utilization in various scenarios, there are still issues that need to be resolved for enabling individualized rehabilitation therapy. On one extreme, a child

with cerebral palsy may have limited arm range of motion, although muscle activity patterns are still present. On the other-hand, a child with development coordination disorder (DCD) may have a viable reaching pattern but requires improvement in hand-eye coordination for successful grasping. Still, children with cognitive disorders, such as autism, may have difficulties comprehending basic interactive scenarios and require a very different type of therapy.

Utilizing the logical fact that animate toys naturally engage children, this research focuses on the design of a robotic therapeutic playmate that will aid children in physical rehabilitation by fusing play and rehabilitation techniques that are both entertaining for the child and effective for upper-arm rehabilitation [13]. The robotic playmate, consistent with prior theory from physical therapy metrics, will incorporate visual queues that will determine the amount of movement that a child makes and his or her consistency with respect to the movement of the robot. Of particular importance within this proposed work are approaches that allow therapists to assess the information obtained during play scenarios, thus requiring that the data is conveniently decipherable.

The following sections introduce background information regarding the development of robotic systems that aid in rehabilitation. First, an overview of pysical ailments with regards to the upper-arm are presented. Next, a brief overview of the current methods that are used by physical therapists to administer upper-arm rehabilitation sessions is given. Finally, a general overview of the current state of the field of therapeutic robotics is provided.

1.1 Upper-Arm Ailments

There are several typical and atypical ailments that affect the upper-arm. A typical ailment can be defined as a decrease in or loss of motor functionality due to scenarios that often affect a large amount of individuals. For instance, sports related injuries,

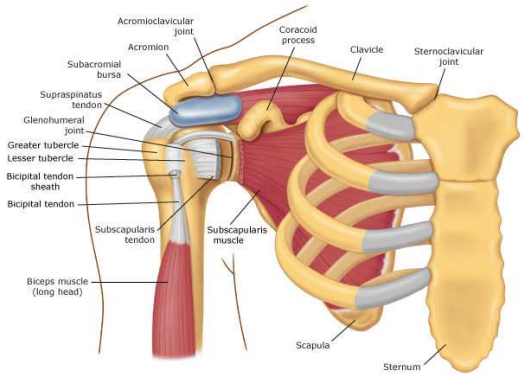


Figure 1: Illustration of the shoulder anatomy [73].

such as a dislocation of the glenohumeral joint (shoulder joint) shown in Figure 1, can cause one to momentarily lose the ability to fully extend, flex, or rotate his or her shoulder and are commonly diagnosed and treated, often resulting in full recovery. Another example of typical ailments are those related to aging, particularly ailments affecting bones and joints such as arthritis. A 2003 National Health Interview Survey projected that 67 million (25%) of adults aged 18 years and older will have doctor-diagnosed arthritis by the year 2030 [22].

Atypical ailments of the upper-arm can also be defined as a decrease in or loss of motor functionality but with regards to individuals that are not indicative of a large portion of society. For instance, the Center for Disease Control (CDC) reports that about 795,000 people in the United States ($\sim 0.26\%$) have a stroke every year [23] and a stroke often affects the motor functions of limbs due to a disconnect between the brain and the specific limb resulting from blood clotting. Another example of an atypical ailment is cerebral palsy. The CDC reports that cerebral palsy is prevalent in 1 in 303 children in the U.S. ($\sim 0.33\%$) and that an estimated 800,000 people in the U.S. ($\sim 0.26\%$) have cerebral palsy [51]. Of the children that are affected, 70 to 80 percent are affected by spasticity [76], which is an involuntary muscle tightness that occurs due to discoordination of their neural motor control and stiffness of their joints, thus making traditional tasks such as throwing difficult. These types of disabilities

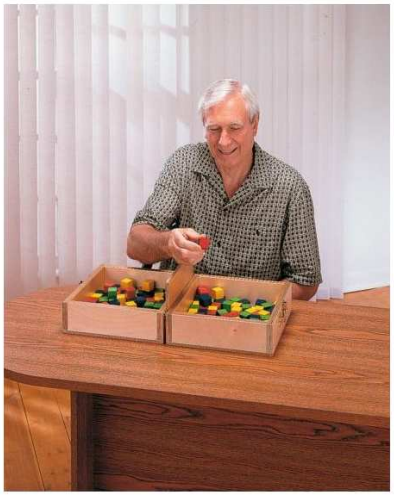


Figure 2: Illustration of a patient peforming the Box and Block Test [71].

have led physicians and researchers to determine various methods for analyzing and aiding patients during therapy in an attempt to provide a higher quality of life.

1.2 Assessments in Physical Therapy

Physical therapists typically use the Action Research Arm Test, the Box and Block Test, and/or the Fugl-Meyer Test to assess upper-arm movement. The Action Research Arm Test [20, 65] assesses the ability to handle small and large objects utilizing a variety of qualitatively rated items and can therefore be considered as an arm-specific measure of activity limitation [82, 114]. The Box and Block Test [33, 67], shown in Figure 2, consists of simply counting the number of blocks that can be transported from one compartment of a box to another compartment within a specified time and is a measure of gross manual dexterity [82], which is defined as an individual's overall ability to utilize the hand to perform tasks. The Fugl-Meyer Test [45] assesses the ability to move the arm and its segments, gives an account of passive joint mobility with an array of qualitatively rated items, and is a measure of impairment [82, 114]. In the late 1990's, a new measurement for child rehabilitation was introduced to the physical therapy field. The Peabody Developmental Motor Scales was designed to measure the fine motor domain of children [42].

1.2.1 The Action Research Arm Test

The Action Research Arm Test (ARA or ARAT) is an observational test used to determine upper limb function. It was first described in 1981 as a modification of an earlier test, the Upper Extremity Function Test (UEFT) [20] and was designed to assess recovery in the upper limb following cortical damage [68]. The test consists of 19 items grouped in subtests (grasp, grip, pinch, and gross arm movement) and performance of each item is rated on a 4-point scale ranging from zero (no movement possible) to three (movement performed normally), with a maximum obtainable score of 57. If a subject scores three on the first, most difficult item of each subtest, it is assumed that they are capable of achieving three on all items of the current subtest; thus, the assessor moves on to the next subtest. If the patient scores less than three, then the second item, the easiest item, is tested. If the patient scores zero on the second item, then it is assumed that he or she is unlikely to achieve a score above zero for the remainder of the items and the assessor moves onto the next subtest. Inter-rater and retest reliability have been shown to be high (ICC > 0.98) in studies involving patients with stroke [68, 108].

1.2.2 The Fugl-Meyer Test

The Fugl-Meyer assessment consists of 155 items and is used to assess impairment, which is any loss or abnormality in psychological, physiological, or anatomical structure or function, in the upper and lower extremities of stroke victims. Utilizing a scoring scale of 226, physical therapists administer a series of tests, which include reflex activity, balance, sensation, position sense, and range of motion (ROM), for evaluating a patient's degree of impairment [38]. The maximum motor performance score is 66 points for the upper extremity, 34 points for the lower extremity, 14, points for balance, 24 points for sensation, and 44 points each for passive joint motion and joint pain. Joint pain is assessed by moving the joint through its available ROM to

assess whether pain occurs at any point in the range [96]. Fugl-Meyer assignes motor function scores to items that assesses motor function alone, with a total possible score of 100 points. Scores are grouped according to the various levels of impairment, which are as follows: <50 points = severe motor impairment, 50-84 points = marked motor impairment, 85-95 points = moderate motor impairment, and 96-99 points = slight motor impairment [44, 96]. The overall reliability of this assessment is high with an intraclass correlation coefficient (ICC) = 0.96 [96]. (The ICC is a statistical assessment describing how strongly units in the same group resemble each other and is typically used when assessing the reliability of a novel method for determining impairment.) The drawback of each of the previously defined assesments is that the scoring scales are based upon the movement expectations of a healthy adult subject.

1.2.3 The Peabody Scale

The lack of a legitimate metric of child motor development led therapists to design scales specifically for assessing the aforementioned group. The Gross Motor Function Measure (GMFM) [95] and the Peabody Developmental Motor Scales (PDMS) [42] are two well-known motor instruments for children, specifically those diagnosed with cerebral palsy (CP). However, the GMFM only measures the gross motor (GM) domain [95]. For measurement of the fine motor (FM) domain, the GMFM is inadequate as an evaluative tool [112]. The opposite is true of the original PDMS.

During the late 1990's, the PDMS was revised to the Peabody Developmental Motor Scales–Second Edition (PDMS-2), with new norms, revised testing materials, more precise scoring criteria, and more information on norm samples [43]. Scores include 1) a Gross Motor Quotient which is a composite of the Reflexes, Stationary, Locomotion, and Object Manipulation subtests, 2) a Fine Motor Quotient, a composite of the Grasping and Visual-Motor Integration subtests, and 3) a Total Quotient, a combination of the gross and motor subtests [43]. The new normative

data on the PDMS-2 was collected on 2,003 children both with and without disabilities. The PDMS-2 is also based on more reliable and valid data sets than the PDMS [43]. Therefore, the PDMS-2 is potentially appropriate for investigating the progress of the gross and fine motor domains for children with and without disabilities because it assesses both GM and FM composites and incorporates both quantitative and qualitative rating criteria [112].

1.2.4 Specific Kinematic Metrics

Since the research presented in this work focuses on non-contact, upper-arm rehabilitation, specifically for the shoulder joint, we present the successful analysis of a typical metric used by physical therapists, namely the Active Range of Motion (AROM). As an extension of AROM, we also quantify the patients' Peak Angular Velocity (PAV) for the purpose of providing a more accurate quantitative analysis and present a method for extracting kinematic data pertaing to reaching. Reaching is typically analyzed by five kinematic characteristics: movement time (MT), total displacement (TD), peak velocity (PV), movement units (MU), and normalized jerk score (NJS) [26, 74]. These types of measurements can coincide with a child's ability to catch and throw objects as well as his or her reaction to environmental events; each of which are within the object manipulation and reflexes subtests of the PDMS-2. Simultaneously, we desire to provide a new method for obtaining this data, namely a way to keep the child subjects motivated during the process.

1.3 Robots in Cognitive and Physical Therapy

Studies involving therapeutic interaction between robots and individuals with cogntive and/or physical disabilities have been of particular interest for several reasons. One particular reason, regarding cognitive development, is that with an understanding of many of these disorders, a solution to a common issue can be determined that will ultimately benefit a broad spectrum of those coping with various disabilities. For

instance, the use of robots for play in children with autism has been of special interest for several reasons. First, the currently accepted way to teach children with autistism involves the use of repetition [34]. Robots are well suited to perform consistent, repetitive actions. Also, it has been shown that children with autism find robots quite engaging and respond favorably to social interactions with them, even when the children typically do not respond socially with humans [35, 70, 97]. The results of these studies may lead researchers to discover cognitive connections between two seemingly different disorders, potentially finding solutions to each.

One of the links between cognitive therapy and physical therapy is the ability to keep the subject engaged throughout the therapeutic process. With repetitive or monotonous conditions over time, performance degradation occurs due to reduced arousal [31]. Utilizing current therapeutic metrics, researchers and physicians are collaborating to discover new, innovative methods for administering therapy sessions for those diagnosed with various physical and cognitive ailments. Robots are a particularly motivating technology because they are concrete, complex, and relate to deep human needs [53]. The following subsections provide a brief overview of some of the current robotic platforms designed for therapeutic use, both cognitively and physically.

1.3.1 Cognitive Therapeutic Robotics

1.3.1.1 Keepon

One robotic platform that has been developed for therapy and play is *Keepon* shown in Figures 3 and 4. Introduced by researchers and physicians in Japan, *Keepon* was designed to engage autistic and non-autistic children (ages 2-4) in playful interaction that was generally initiated and directed by the child (i.e. without any experimental setting or instruction) [58]. A point of emphasis for the *Keepon* design was to keep it as simple as possible in order to abstain from overwhelming and/or frightening the children, a principle that all researchers working with therapeutic robots should keep

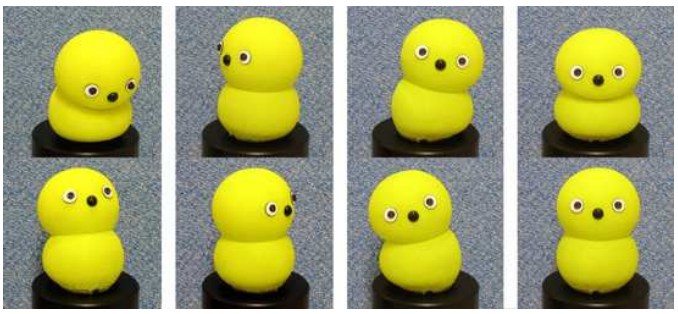


Figure 3: Illustration of Keepon in various movement positions [58].

Figure 4: Illustration of the interior machinery of Keepon [58].

in mind. Therefore, the simple structure and function of the robot contained two color CCD cameras as eyes, a microphone as a nose, a small gimbal and four wires, by which the body is manipulated like a marionette, four motors and two circuit boards (an SH2-based PID controller and a motor driver). Since the body is made of silicone rubber and its inside is relatively hollow, *Keepon's* head and belly deform whenever it changes posture and when people touch it [58].

The resulting information, obtained from a year and a half of data collection, indicated that *Keepon's* simple appearance and predictable responses gave autistic children a playful and relaxed mood, in which they spontaneously engaged in dyadic play with *Keepon*. This in turn expanded into interpersonal communication where *Keepon* worked as the pivot of triadic play with adults or other children [58].

(a) (b) (c)

Figure 5: Illustration of KASPAR demonstrating facial expressions [6].

1.3.1.2 KASPAR

Facial expressions are a widely used feedback mechanism for human-human interaction. By utilizing visual cues that are commonly understood among humans, one can determine the type of emotion that is currently being displayed by a counterpart during an interactive scenario. Furthermore, it is quite common that individuals with severe disabilities, such as autism and cerebral palsy, have difficulty communicating via speech and/or movement of the limbs. Thus, communication through facial expressions holds merit.

One robotic system that uses this approach to aid in therapeutic sessions is KASPAR (Kinesics And Synchronisation in Personal Assistant Robotics). KASPAR is a child-sized robot which acts as a platform for HRI studies, using expressions and gestures to communicate with a human interaction partner. The goal of this robotic system is to induce optimal realism for rich interaction while avoiding the uncanny valley [6]. The uncanny valley refers to a person's sense of unease and discomfort when they look at increasingly realistic virtual humans [72]. As a machine acquires greater similarity to a human, it becomes more emotionally appealing to the observer. However, when it becomes disconcertingly close to a human there is a very strong drop in believability and comfort, before finally achieving full humanity and eliciting positive reactions once more [9].

The original robotic design consisted of a 8-DOF head and a static body, shown in

Figure 6: The CosmoBot^{TM} robot (left), Mission Control^{TM} (interface board device, center right), the Therapist Modul software (interface shown on module, back right), and plug-in gestural interfaces: a joystick, and wearable head and arm sensors (front center) [11].

Figure 5, while the modified design includes two 6-DOF arms. KASPAR can be used to study a variety of research issues relevant to HRI such as interaction dynamics, gesture creation and recognition, joint attention, communication through imitation, and the use of expressions [6].

1.3.1.3 CosmoBot

Another robotic system that has been designed and developed for therapy, education, and play is the CosmoBot^{TM}. Children interact with the CosmoBot^{TM}, controlling the robot's movements and audio output, using a variety of gestural sensors and speech recognition, while actively targeting their therapy goals [11]. The system, shown in Figure 6, consists of the physical robot, an interface board, interface software, and plug-in gestural interfaces (joystick, head sensor, and arm sensor). Figure 7 illustrates the attentiveness of a child with cerebral palsy to the robotic playmate.

Researchers recruited six children who were already receiving outpatient physical and occupational therapy services with a physical therapist. More specifically, the subjects ranged in age from 4 – 10 years old, and each subject had a diagnosis of cerebral palsy and received physical and/or occupational therapy for treatment of

Figure 7: Illustrations of children diagnosed with cerebral palsy interacting with the Cosmobot, a telerehabilitation robot. [11]

(a) Normal appearance (b) Undressed to show parts

Figure 8: Illustration of Robota [92].

upper extremity motor deficits. The physical therapist provided feedback regarding the system, noting that it was easy to use and that it provided a motivating factor for client participation [11].

1.3.1.4 Robota

The Aurora project is an ongoing research effort to aid in the therapy and education of children with autism. In this project, scientists utilize a humanoid robotic doll, named Robota, to engage children with this disability in imitation games. Robota, shown in Figure 8, is connected through a serial link to a PC and can use speech synthesis, speech processing, and video processing of data from a quick-cam camera.

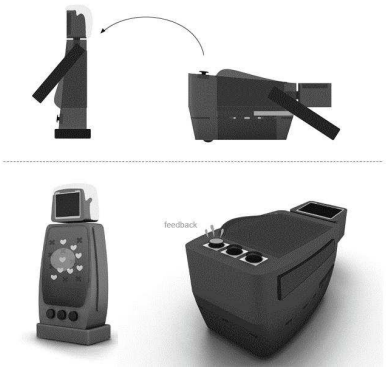


Figure 9: Illustration of the IROMEC robotic toy in its horizontal and vertical configurations [80].

Using its motion tracking system, Robota can copy upward movements of the user's arms, and sideways movements of the user's head when the user sits very still and close to the robot, looking straight at it, engaging in turn-taking and imitation games with the robot [92].

The initial results, however, were inconclusive, and a number of drawbacks of the original setup were identified. Consequently, the researchers decided to revamp the procedure by using a much more unconstrained setup, pursuing a longitudinal study, and reducing the intervention of caregivers. The researchers focused their analysis on four specific areas: Eye Gaze, Touch, Imitation, and Proximity, each with regards to the child's relation to Robota [92]. Quantitative and qualitative results of this project indicated that the level of child engagement to the robot was proportional to the allotted time given for interaction and that the children actively sought to involve the investigator and caregivers in their interaction with Robota when given sufficient time to become accustomed to the system.

1.3.2 IROMEC Robotic Social Mediator

The IROMEC project (Interactive Robotic Social Mediators as Companions) recognizes the important role of play in child development and targets children with

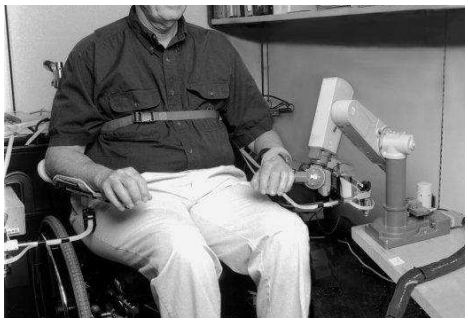


Figure 10: Illustration of the Mirror-Image Motion Enabler

autism, children with mild mental retardation, and children with severe motor impairments [80]. The IROMEC robotic toy, shown in Figure 9 can play a therapeutic role for children with autism by encouraging the development of communication, motor, cognitive, sensory, and social interaction skills towards joyful experiences in interactive play [80]. Since the behavior of children with autism can vary from one patient to the next, the IROMEC robot was designed with the capabilities of being tailored to the user's needs by utilizing various input/output devices, such as dynamic screens, buttons, and wireless switches.

The researchers in this particular project used 10 play scenarios for robot assisted play and robotic mediator in order to teach children valuable communication and turn-taking skills. The IROMEC robotic toy specifically aimed to assist in the development of seven different functions: global intellectual function, memory functions, higher-level cognitive functions, copying, attention functions, energy and drive functions, and undertaking of task [80]. Utlizing play scenarios such as imitation play and sensory reward, the robot adapted itself to the child, helping him or her to explore and reaching increasingly complex objectives needed for social interaction dynamics [80]. Results of various studies using the IROMEC robotic toy are said to be forthcoming in future publications.

1.3.3 Physical Therapeutic Robotics

1.3.3.1 Mirror-Image Motion Enabler

The Mirror-Image Motion Enabler or MIME [15], shown in Figure 10, was designed to aid in the physical rehabilitation of stroke victims. This platform functioned by strapping the subject into a wheel-chair that had two mobile arm supports modified to limit arm movement to the horizontal plane and a six degree of freedom (DOF) robot arm (PUMA-260) that applied forces and torques to the paretic forearm through one of the arm supports. A 6-axis transducer measured the force/torque interaction between the robot and the paretic limb, and movements were produced in preprogrammed forearm position and orientation trajectories or by a position feedback control system that slaved the robot to the movements of the contralateral (normal) limb [15]. The initial results of testing 13 hemiplegic male subjects indicated that MIME successfully assisted the paretic arm movements of 12 out of the 13 individuals. Futher testing of 21 subjects (11 within the robotic group and 10 within the control group) indicated that robot-assisted therapy may have advantages over conventional NeuroDevelopmental Therapy-based techniques as most subjects tested exhibited improvements in the Fugl-Meyer assessment of motor function [15].

1.3.3.2 Hands-Off Mobile Platform

Another robotic platform with the capability of providing an assessment of a patient's physical attributes was introduced by Gockley and Matarić in 2006. The researchers in this project attached a laser range finder to a Pioneer-2DX robotic platform, shown in Figure 11, that was used to track and locate a patient wearing a reflector on his or her leg. By tracking the patient's movement, the system could determine the patient's arm position during the therapeutic session. The researchers also used the system as a way of encouraging the patient to continue the rehabilitation session through positive reinforcement, which was provided via a series of beeps and movements based

Figure 11: Illustration of the Pioneer-2DX [47].

on the patient's movements. As a control variable, they varied the robot's level of engagement (or amount of movement) and they reported the patients' level of engagement and comfort with the robot [47]. Testing of 11 individuals indicated that participants tended to exercise longer with the robot when they perceived that it was engaged with them, supporting the use of robots in physical therapy for monitoring efforts and progress.

1.3.3.3 ARMin

The ARMin rehabilitation robot, shown in Figure 12 is part of an on-going research project to improve upper extremity function after spinal cord injury, stroke, or neurological diseases [91]. The first version of the robot (ARMin I) consisted of four DOF actuating the shoulder in three dimensions and flexing/extending the elbow. During the years of 2003 to 2006, therapist and researchers tested the initial version on several patients at the University Hospital Balgrist Zürich. Each patient's affected arm was connected to the robotic platform by an end-effector and he or she operated the device in three different modes: passive mobilization, active game-supported arm therapy, and active training of Activities of Daily Living (ADL) tasks.

The second version (ARMin II) included a complete exoskeletal structure with two

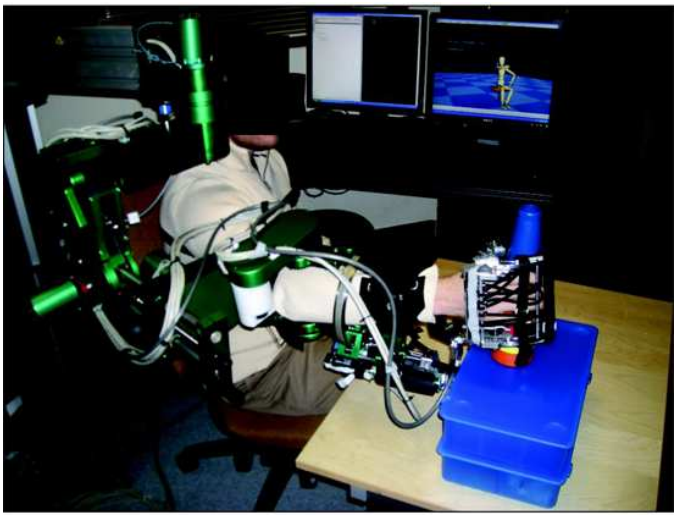


Figure 12: Illustration of the ARMin III robotic platform used for upper-arm rehabilitation [12].

additional degrees of freedom (six DOF altogether) also allowing pronation/supination of the lower arm and wrist flexion/extension [91] and focused heavily on ensuring proper movement of the shoulder in the vertical direction. The third version (ARMin III) was improved by adding a seventh DOF that enabled the opening and closing of the hand. New ADL tasks such as setting a table, cooking, filling a cup, opening a door, personal hygiene, using a vending machine, and playing a piano [91] were added to create a more robust testing environment that would engage and challenge the subjects. Finally, version four (ARMin IV) was equipped with three high-precision force sensors placed close to the cuffs of the robot allowing the recording of 18 degree-of-freedom forces and torques interacting between the robot and patient [91].

Pilot testing of the robot showed moderate but significant improvements in all subjects with regards to the Fugl-Meyer Assessment pertaining to the upper-limb. Further testing showed improvements in three out of four subjects pertaining to the Wolf Motor Function Test (WMFT) [113, 117]. A larger, more detailed, publication of the results are expected early this year (2012).

Figure 13: Illustration of the intelligent Pneumatic Arm Movement [52].

1.3.3.4 intelligent Pneumatic Arm Movement

iPAM (intelligent Pneumatic Arm Movement) is a dual robotic system that aims to assist in the recovery of upper-limb movement in people with all severities of motor impairment after stroke [52]. Specifically geared towards reaching exercises, the robot consists of three active pneumatically powered rotational joints that are connected to the patient's upper limb via a passive orthosis that allows three passive rotational DOF [52]. The unit allows for six DOF of the upper-limb (five DOF at the shoulder and one at the elbow), rotation measurements of the six active robot joints using rotary sensors, and recording of the relative forces between the robot and the human's limb using force/torque transducers. The design allowed assisted exercises to be pre-recorded by a physical therapist in 3D exercise workspace and can be replayed in active assistance mode during the rehabilitation of a patient.

A pilot study for 16 post-stroke individuals, each performing six to eleven exercises, indicated promising results for analyzing reaching movements. Position and force data allowed the researchers and therapist to determine the effectiveness of the robotic unit as well as future treatment for the subjects. Patient feedback revealed that the robotic platform needed improvement in the area of comfort and patient attachment to the device. The feedback that was gathered was used to refine the device for a second model which incorporated a custom made frame and padded back support, a docking

Figure 14: Illustration of the SAR (Bandit) platform used to motivate physical exercise for older adults [39].

station for the patient chair, and improved performance and reliability of the system by refining the valves and cylinders. More testing of the system in a larger pool of subjects is expected in the immediate future.

1.3.3.5 Bandit

Most recently, Fasola and Matarić conducted a study using a socially assistive robot (SAR) to motivate physical exercise for older adults [39]. The researchers specifically designed a robotic platform to act as an administrator and active participant in a health-related activity. The researchers noted that the social interaction between the robot and user is not only useful for maintaining user engagement, but is also instrumental in achieving the physical exercise task [39]. The SAR, shown in Figure 14, was capable of providing positive verbal feedback that was hypothesized to induce intrinsic motivation with regards to keeping the subject engaged during each exercise. The SAR was also capable of physically interacting with each subject by demonstrating or mimicking exercises in various game scenarios while visually monitoring the user's performance based upon an arm pose recognition algorithm.

The SAR, named Bandit, was a biomimetic anthropomorphic robot platform that

consisted of a humanoid torso mounted on a MobileRobots Pioneer 2DX mobile base [39]. It consisted of 19 degrees of freedom (DOF) and used a standard USB camera located at the waist to capture the user's arm movements. Its speech was generated via the NeoSpeech text-to-speech engine, and speakers were used to output the voice.

The researchers studied two aspects as it related to physical exercise and engagement: 1. Praise and relational discourse effects and 2. User choice and self-determination. The first study used different coaching styles coupled with a post-test questionnaire to assess the relational value of the robot and objective measures to assess the user performance of each interaction. The results of the first study indicated that subjects showed a stronger user preference for the relational robot over the non-relation robot, demonstrating the positive effects of praise and relational discourse in a healthcare task-oriented human-robot interaction scenario [39]. The researchers noted that the subjects also consistently engaged in physical exercise throughout the interaction as indicated by the objective measures obtained from the vision algorithm. The second study used Choice and No-Choice conditions to test user preferences regarding choice of activity. A post-test questionnaire revealed no clear preference for one condition over the other, but indicated that the SAR was helpful and attributed to the exercise [39].

1.3.4 Moving Forward

These systems represent the state-of-the-art in robotic platforms used in therapy. While these robotic systems are among the first to make significant progress towards aiding individuals with severe disabilities, there is still much that needs to be solved. First, the robots in the cognitive therapy group do not give a physical assessment of the individuals' motor skills. Even though children with disabilities may never have full use of certain body parts, significant improvements can be made in hopes that these individuals will be able to lead healthy, self-contained lifestyles. Also, of

the systems that do incorporate a physical assessment of the individual, those in the physical therapeutic group, only two (the hands-off mobile platform and the SAR) attempt to assess the cognition of the patient during the therapy while assessing the physical movement. Still, these specific platforms lack the physical appeal, cost-effectiveness, and robust objective analysis needed to be incorporated into a broad spectrum of rehabilitation therapy.

With the introduction of these robotic systems, there is research showing improved cognitive development and, separately, contact assistive robotics that incorporate objective kinematic measures. By bridging the gap between these two research areas, a system can be designed that allows a patient to move freely while keeping them engaged and motivated during a continuous rehabilitation session as it calculates the patient's kinematic measures as they pertain to specific exercises. Furthermore, since the progression of robotics has been to achieve full autonomy, it is ideal to produce systems with the capabilities of making intelligent decisions and performing actions without the aid of a separate system, be it human or computer. This work is a progression towards achieving these goals.

CHAPTER II

IMAGE SEGMENTATION

Often, researchers utilize a motion capture system to accurately track a subject's movements during exercise and analyze the complete movement after the data has been stored [16, 25, 74]. However, when considering in-home treatment, it is impractical to administer this method of tracking. As an alternative, robust computer vision algorithms can enable segmentation and tracking of the patient's movements while simultaneously ensuring the feasibility of incorporating a cost-effective and space-efficient device (e.g. a camera). With the introduction of the Microsoft Kinect®, obtaining depth information for a more detailed segmentation is relatively straightforward. The Kinect, introduced by PrimeSense for gaming with the Xbox 360, is a device that has two cameras and one Infrared (IR) projector that enables it to obtain depth information. Many researchers in the field of computer vision have migrated to utilizing this groundbreaking technology for several applications [102, 116]. Concurrently, this research uses the Kinect's IR Projector in order to extract subjects' movements in three-dimensions (3-D).

In the initial approach of quantifying the movements of human subjects during a physical therapy session, it was desirable to choose a particular method that would capture a subject's movements while simultaneously removing the static background. In other words, it is desirable isolate the human in each image to ensure that the motions of the subject are the sole focus. Among the many well-known processes that can be used to achieve this result, we focused on two that could be utilized within our framework. The next sections describe both methods discuss the implementation and benefits of each. More specifically, we show how each process can be applied to

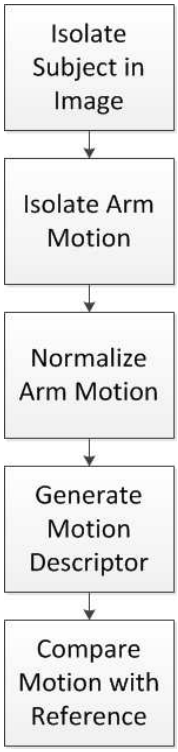


Figure 15: Chronological flowchart of the overall image segmentation process.

a subject's upper-limb exercise data with regards to isolating the limb in question. Figure 15 is a chronological flowchart of the overall process.

2.1 Segmentation of the Image Background

The initial step in quantifying the subject's movement is to segment a video sequence into individual images that contain pertinent information as it relates to the overall representation of recent movement [14]. One common technique for attaining 3-D information from a particular movement is to recover the pose of the person at each time instant using a 3-D model [7]. This generally requires a strong segmentation of foreground/background and also of individual body parts to aid the model alignment process. Furthermore, background uniformity is typically a necessity when processing images for motion estimation [48, 89]. However, in this work, it is desirable to enable human-robot interaction and data collection in real-time rather than require the subject to wait during an initiation process.

(a) Subject's starting position. (b) Subject's ending position. (c) MHI of sequence.

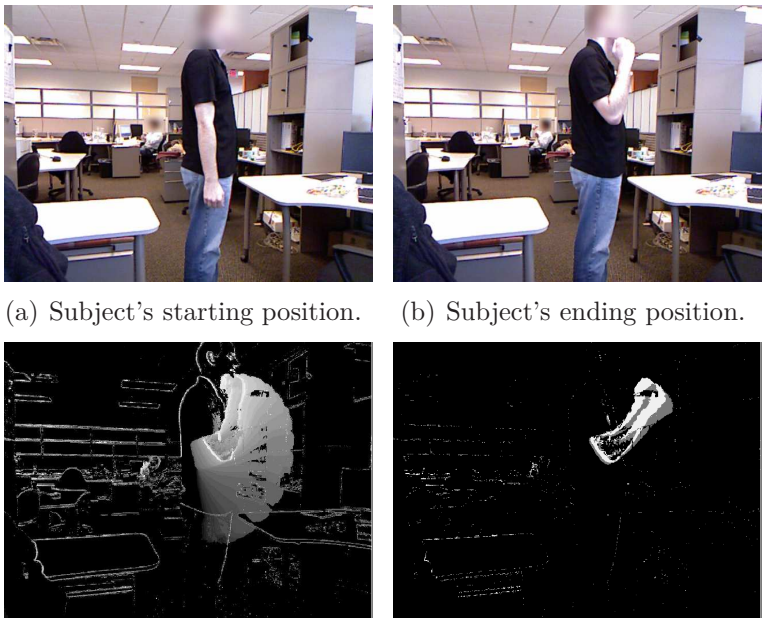
Figure 16: Motion History Image of the left shoulder abduction.

2.2 Motion History Imaging Algorithm

Since the purpose of this work is to analyze the movement of specific body parts, similar to [18], the first algorithmic approach uses temporal templates. While some algorithms utilize sequences of static configurations, which require recognition and segmentation of the person [83], here, a Motion History Image (MHI) is used to represent how motion in the image is moving. This essentially allows real-time processing of the input data. In a MHI, pixel intensity, H_τ, is a function of the temporal history of motion at a point (x, y) point in physical space [7]. The goal of the MHI process is to monitor the input for changes in pixel intensity and track those changes over time. Similar to Bobick and Davis, a replacement and decay operator is used, as shown in Equation (1), to obtain the MHIs:

$$H_\tau(x, y, t) = \begin{cases} \tau & \textit{if } D(x,y,t)=1 \\ max(0, H_\tau(x, y, t-1) - 1) & \textit{otherwise} \end{cases} \quad (1)$$

where D is a binary image sequence indicating regions of motion, x and y are the horizontal and vertical directions in the image, respectively, t is the current time step, and τ is the current intensity value [7]. The result, as illustrated in Figure 16, is a scalar-valued image where more recently moving pixels are brighter in intensity. Figure 17 illustrates the results of varying the decay factor on the same dataset.



(a) Subject's starting position. (b) Subject's ending position.

(c) MHI of sequence, illustrating full motion. (d) MHI of sequence, illustrating decay operator equal to $\frac{\tau}{2}$.

Figure 17: Motion History Image of the right elbow flexion.

2.3 Multimodal Mean Algorithm

The Multimodal Mean (MM) is an algorithm which models each background pixel as a set of average possible pixel values. In background subtraction, each pixel I_t in the current frame is compared to each of the background pixel means to determine whether it is within a predefined threshold. The background model for a given pixel is a set of K mean pixel representations called cells. Given each cell contains three mean color component values, an image pixel I_t is a background pixel if each of its color components is within a predefined threshold for that color component E_x of one of the background means [2].

Each background cell B_i is represented as three running sums for each color component $S_{i,t.x}$ and a count $C_{i,t}$ of how many times a matching pixel has been observed in t frames. At any given frame t, the mean color component value is then computed as $\mu_{i,t.x} = S_{i,t.x}/C_{i,t}$. More specifically, I_t is a background pixel if a cell B_i can be found whose mean for each color component x matches within E_x of the corresponding color

component of I_t:

$$\left(\bigwedge_x |I_{t.x} - \mu_{i,t-1.x}| \leq E_x \right) \wedge (C_{i,t-1} > T_{FG}), \tag{2}$$

where T_{FG} is a small threshold representing the number of times a pixel value can be seen and still considered to be foreground [2].

When a pixel I_t matches a cell B_i , the background model is updated by adding each color component to the corresponding running sum $S_{i,t.x}$ and incrementing the count $C_{i,t}$. In addition all cells can be periodically decimated by halving both the sum and the count every d (the decimation rate) frames in order to enable long-term adaptation of the background model. This decimation can prevent background components from permanently dominating the model, thus allowing the background model to adapt to the appearance of newer stationary objects or newly revealed parts of the background. It also plays a secondary role in the embedded implementation in preventing counts from overflowing their limited storage within memory. More specifically, when I_t matches a cell B_i , the cell is updated as follows:

$$S_{i,t.x} = (S_{i,t-1.x} + I_{t.x})/2^b \tag{3}$$

$$C_{i,t} = (C_{i,t} + 1)/2^b \tag{4}$$

where b = 1 if t mod d = 0, and b = 0, otherwise [2].

When a pixel I_t does not match cells at that pixel position, it is designated as foreground. In addition, a new background cell is created to allow new scene elements to be incorporated into the background. If there are already K background cells, a cell is selected to be replaced based on the cell's overall count $C_{i,t}$ and a recency count $R_{i,t}$ which measures how often the background cell's mean matched a pixel in a recent window of frames. A sliding window is approximated by maintaining a pair of counts

$(r_{i,t}, s_{i,t})$ in each cell B_i . The first $r_{i,t}$, starts at 0, is incremented whenever B_i is matched, and is reset every w frames. The second $s_{i,t}$, simply holds the maximum value of $r_{i,t}$ computed in the previous window [2]:

$$r_{i,t} = \begin{cases} 0 & \text{if t mod w} = 0 \\ r_{i,t-1} + 1 & \text{if } B_i \text{ matches } I_t \text{ and t mod w} \neq 0 \end{cases} \tag{5}$$

$$s_{i,t} = \begin{cases} r_{i,t-1} & \text{if t mod w} = 0 \\ s_{i,t-1} & \text{if otherwise.} \end{cases} \tag{6}$$

Recency $R_{i,t} = r_{i,t} + s_{i,t}$ provides a measure of how often a pixel matching cell B_i was observed within a recent window. The $s_{i,t}$ component allows information to be carried over across windows so that recency information is not completely lost at window transitions. When a new cell is created and added to a background model that already has K cells, the cell to be replaced is selected from the subset of cells containing oldest cells, i.e., cells whose recency $R_{i,t} < w/K$. From this set, the cell with the minimum overall count $C_{i,t}$ is selected for replacement. The cell with lowest $C_{i,t}$ is replaced in the rare event that all cells have a recency count $R_{i,t} > w/K$ [2].

2.4 Comparison of Background Image Segmentation Algorithms

The MHI algorithm is solely based upon frame differencing D and the decay operator τ, whereas the MM algorithm combines various methods of thresholding with running sums and sliding window approximations. As such, it is logical that the MM algorithm will take longer to process images than the MHI algorithm (testing revealed 2x the time). However, the MM algorithm is much more accurate at achieving the desired segmentation than the MHI algorithm.

Both algorithms work well for standard red, green, blue (RGB) and grayscale images. For this work, however, the IR Projector, which essentially is an IR laser that passes through a diffraction grating and converts into many IR pixels, and IR

(a) Original (b) MM (c) MHI

Figure 18: Result of MM and MHI algorithms as standalone methods for segmenting the IR projected image.

camera pair on the Microsoft Kinect® was used. Each pixel is unique to the IR camera, which determines the coordinates of the pixel, thus allowing realtime 3-D information. The downfall is that the IR projector produces quite a bit of noise due to lighting conditions, low resolution, and other factors, which causes each image segmenation process to produce noisy segmentations. Ultimately, this would cause an improper recognition of the subject's arm locations throughout the excerise. The results of each standalone process is shown in Figure 18.

One option would be to segment the Kinect's RGB images using the MM or MHI algorithm and map the segmented pixels to corresponding pixels in the depth images. However, after testing this procedure, it was realized that often times the images produced by the Kinect's RGB camera were not properly aligned with those produced by the IR Projector and IR camera. Therefore, one would need to manually determine the proper alignment between the two cameras with each collection of data. Since this is undesirable, as the robotics world should be moving towards more autonomy, another solution was warranted. As a result, it was determined that it was best to use *a priori* information. If each person that interacts with the robot is placed at a known distance, segmenting the person from the background is straightforward. By removing all data that is outside of the distance value produced by the IR projector, representing the location of the subject, a less noisy silhouette is produced, see Figure 19.

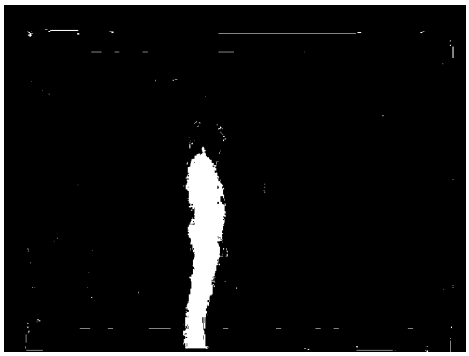
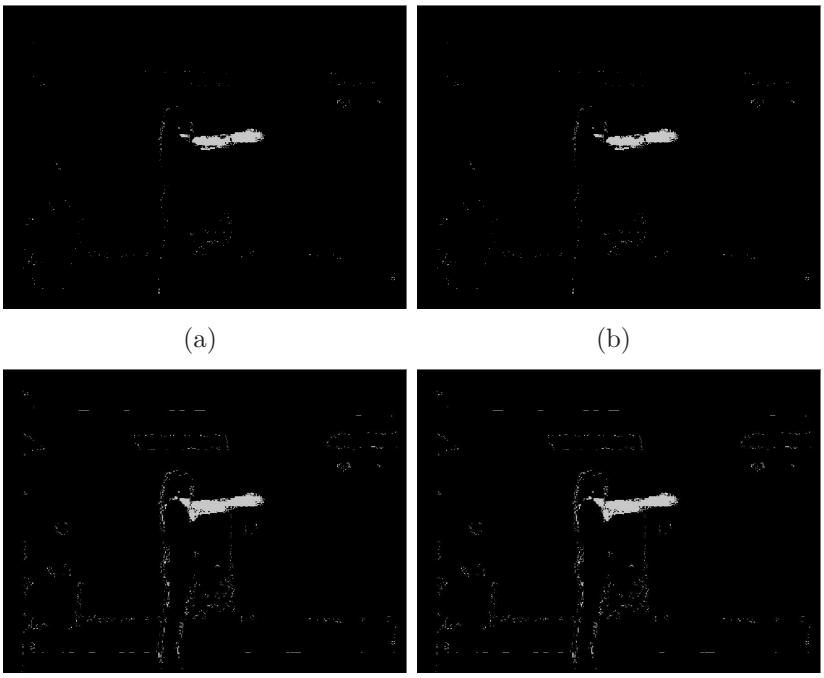


Figure 19: Silhouette of subject given a known distance from the IR projector.

However, as seen in the figure, there are still a few artifacts produced by the IR Projector and IR camera pair. The noisy silhouette (coupled with subject movement) was tested with the MM algorithm using different values for the maximum component difference threshold (MCDTH), which is the pixel difference threshold between frames, as well as the cell threshold (CTH), which is the number of cells in a given window used to calculate the recency value $R_{i,t}$. Figure 20 illustrates the results of the various trials. Figure 21 illustrates the results of varying the replacement and decay operator, $\Delta\tau$, in the MHI algorithm with the same set of images.

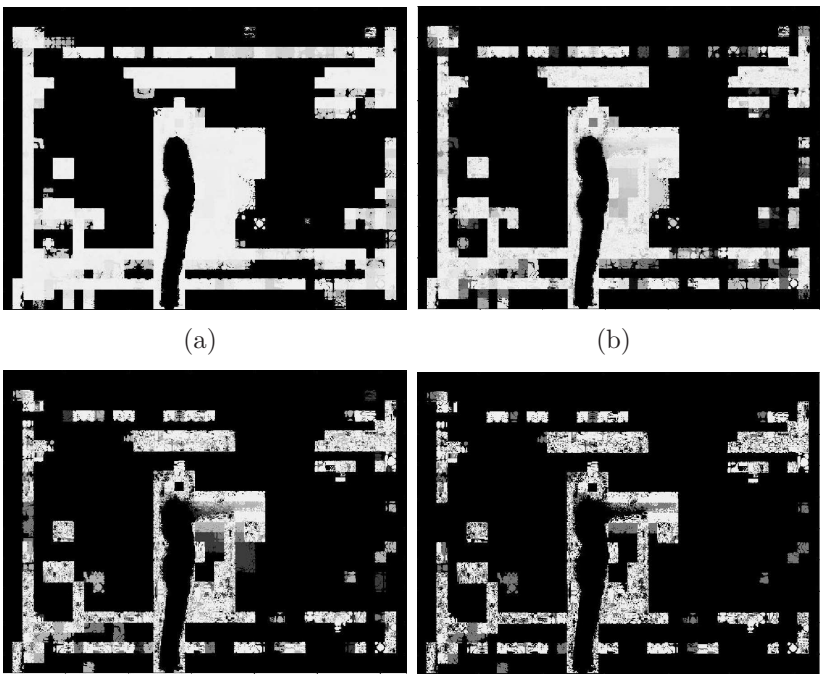
It is immediately evident that the resulting images for the MM algorithm are better than those produced by the MHI algorithm when coupling with a distance segmented IR projected image. However, each of the resulting images still contained unwanted artifacts, and, if not chosen properly, the MCDTH and CTH values may diminish the desired arm segmentation. Thus, it was determined that a simple filter would be beneficial in order to remove excess noise prior to calculating the patient's range of motion. Initially, the reseachers utilized a median filter to remove the noise within the image. A median filter is a sliding-window spatial filter that replaces the center pixel value in the window with the median of all the pixel values in the window, and it can be of any central symmetric shape, a round disc, a square, a rectangle, or a cross. Equation (7) shows a mathematical calculation for each of the central pixels in the sliding windows, where $I_{i,j}$ is a central pixel, $I_{k,l}$ is each pixel within the



(a) (b)

(c) (d)

Figure 20: MM results for different values of MCDTH and CTH: (a) MCDTH = 0.001, CTH = 4, (b) MCDTH = 1.000, CTH = 4, (c) MCDTH = 0.001, CTH = 10, (d) MCDTH = 1, CTH = 10



(a) (b)

(c) (d)

Figure 21: MHI results for different values of $\Delta\tau$: (a) $\Delta\tau = 1$, (b) $\Delta\tau = 10$, (c) $\Delta\tau = 60$, (d) $\Delta\tau = 120$

Figure 22: Median filtered image.

window, m and n are the row and column respectively, and N is the total number of pixels in the window. For this approach, a 16x16 square window was used. Figure 22 illustrates the use of a median filter on image data.

$$I_{i,j} = \frac{\sum_{k=1}^{m}\sum_{l=1}^{n} I_{k,l}}{N} \tag{7}$$

The major downfall of using a median filter is the time complexity that it takes to determine the median within an unsorted list of pixels for a certain number of windows ($O(mn)$). This induced a search for a more efficient method that yielded similar results. By performing a Gaussian blur coupled with Suzuki's well-known process for border following [103] to extract and group the contour(s) within each image, the largest contour (based on area) representing the subject's arm movement can be determined. Suzuki's process basically determines the borders of a group of pixels by utilizing parent bordering and a predefined marking policy that marks each pixel along the border. A Gaussian blur is a type of image-blurring filter that uses a Gaussian function for calculating the transformation to apply to each pixel in the image. The equation of a Gaussian function in one dimension is

$$G(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}} \tag{8}$$

where x is the distance from the origin in the horizontal axis, y is the distance from the

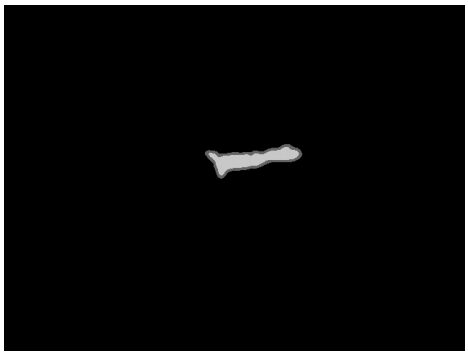


Figure 23: Resulting image after Gaussian blur and contour size threshold.

origin in the vertical axis, and σ is the standard deviation of the Gaussian distribution. In two dimensions, it is the product of two such Gaussians, one in each dimension:

$$G(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \tag{9}$$

When applied in two dimensions, this formula produces a surface whose contours are concentric circles with a Gaussian distribution from the center point. Values from this distribution are used to build a convolution matrix which is applied to the original image. Each pixel's new value is set to a weighted average of that pixel's neighborhood. The original pixel's value receives the heaviest weight (having the highest Gaussian value) and neighboring pixels receive smaller weights as their distance to the original pixel increases [98]. The resulting image is shown in Figure 23.

Furthur speculation lead to testing as to whether or not using the Gaussian blur had a significant effect on the output of the image after using the contour threshold. Since the threshold only returns the largest contour, one only needs to utilize the portion immediately following the MM segmentation in order to achieve the desired affect of isolating the subject's arm in each image. The resulting image is shown in Figure 24; although not a smooth as the Gaussian blur image, this approach, nonetheless, yields the desired result while removing a step in the overall process.

Figure 24: Resulting image after contour size threshold, but without Gaussian blur.

2.5 Finding World Coordinates

Once the subject's upper-limb has been properly segmented, the next step is to determine the correlating world coordinates of each pixel representing the arm's location within the image. Classic computer vision algorithms determine the depth of specific objects using a RGB stereo camera pair. The designer's of the Microsoft Kinect (PrimeSense) have incorporated this aspect using a different approach, namely two cameras (one IR and one RGB) coupled with a laser-based projector. Since the IR Projector and IR camera pair only return the depth information of the segmented arm, traditional computer vision techniques have to be applied in order to extract the arm's world coordinates.

For a pinhole camera model, a 3D scene point P with world coordinates (X,Y,Z) is projected to a 3D point Q in the virtual image plane. By simply rescaling, it is noted that the coordinates for Q with respect to P are:

$$Q = \left(f\frac{X}{Z}, f\frac{Y}{Z}, f \right) \tag{10}$$

where f is the focal length of the camera. Hence, the 2D coordinates in the virtual image are given by:

$$(u, v) = \left(f\frac{X}{Z}, f\frac{Y}{Z} \right) \tag{11}$$

The same basic concept applies for a stereo camera pair with the exception that the cameras are shifted along the x-axis. With this in mind, the equation needs to compensate for that shift. Thus, if the model is taken with respect to a camera on the left, the 2D coordinates of the camera on the left will remain the same as in Equation 11 (now denoted by u_L and v_L), and the 2D coordinates for the camera on the right will become:

$$(u_R, v_R) = \left(f\frac{X-B}{Z}, f\frac{Y}{Z} \right) \tag{12}$$

where B is the baseline distance between the two cameras. Thus, the disparity d between the two cameras is given by:

$$d = (u_L - u_R) = f\frac{B}{Z} \tag{13}$$

Solving for the depth Z, gives:

$$Z = f\frac{B}{d} \tag{14}$$

The Kinect disparity is related to a normalized disparity by the relation

$$d = \frac{1}{8}(d_{off} - k_d) \tag{15}$$

where d is a normalized disparity, k_d is the Kinect disparity, and d_{off} is an offset value particular to a given Kinect device. The factor $\frac{1}{8}$ appears because the values of k_d are in $\frac{1}{8}$ pixel units. The value for B is always about 7.5 cm, which is consistent with the measured distance between the IR and projector lenses, and d_{off} is typically around 1090 [57]. Utilizing this information, the calculations for finding the (X, Y, Z) coordinates of the arm become:

$$X = \frac{(u-k_{xd})Z}{f}$$
$$Y = \frac{(v-k_{yd})Z}{f} \tag{16}$$
$$Z = Z$$

where k_{xd} and k_{yd} are the Kinect disparities in the x and y directions respectively. These values are calculated via the calibration process found in [57].

2.6 Principal Component Analysis

Once the world coordinates of the segmented arm have been determined, the next step is to make certain that all movements are planar with respect to the camera. Adding this component makes it simpler to calculate the therapeutic metrics described later by diminishing inaccuracies resulting from typical human characteristics. That is to say, human movements will undoubtedly have slight variations in the depth plane. Since physical therapists typically analyze patient kinematics with respect to the saggital plane, it is advantageous for this approach to correlate with that point of view. Thus, the rotation to planar view needs to eliminate the aforementioned variations.

The rotation to planar view is achieved by performing a Principal Component Analysis (PCA) on the set of images and projecting the pixels of each image onto the new plane using the resulting eigenvectors. A PCA is mathematically defined as an orthogonal linear transformation that transforms a set of data to a new coordinate system such that the greatest variance by any projection of the data comes to lie on the first coordinate (called the first principal component), the second greatest variance on the second coordinate, and so on [54]. PCA can be achieved either by an eigenvalue decomposition of a data covariance matrix or singular value decomposition of a data matrix. The eigenvalue decomposition of a square symmetric matrix, $\mathbf{A}$, is given by:

$$\mathbf{A} = \mathbf{P}\mathbf{\Lambda}\mathbf{P}^{\mathbf{T}} \tag{17}$$

where $\mathbf{P}$ is an orthogonal matrix with columns $\mathbf{X_1}$,...,$\mathbf{X_n}$ which are linearly independent eigenvectors of $\mathbf{A}$ and $\mathbf{\Lambda}$ is a diagonal matrix that contains the corresponding eigenvalues of $\mathbf{A}$ across its diagonal. A more detailed explanation of the eigenvalue decomposition is given in [86].

Given a data matrix, $\mathbf{X}$, its singular value decomposition is defined as:

$$\mathbf{X} = \mathbf{W}\mathbf{\Sigma}\mathbf{V}^{\mathbf{T}} \tag{18}$$

where $\mathbf{W}$ is the m x n matrix of eigenvectors of $\mathbf{X}\mathbf{X}^{\mathbf{T}}$, $\mathbf{\Sigma}$ is an m x n rectangular diagonal matrix with nonnegative real numbers on the diagonal, and matrix $\mathbf{V}$ is n x n the matrix of eigenvectors of $\mathbf{X}^{\mathbf{T}}\mathbf{X}$. The PCA transformation that preserves dimensionality is then given by:

$$\begin{aligned} \mathbf{Y}^{\mathbf{T}} &= \mathbf{X}^{\mathbf{T}}\mathbf{W} \\ &= \mathbf{V}\mathbf{\Sigma}^{\mathbf{T}} \end{aligned} \tag{19}$$

Since $\mathbf{W}$ (by definition of the singular value decomposition of a real matrix) is an orthogonal matrix, each row of $\mathbf{Y}^{\mathbf{T}}$ is simply a rotation of the corresponding row of $\mathbf{X}^{\mathbf{T}}$. The first column of $\mathbf{Y}^{\mathbf{T}}$ is made up of the "scores" of the cases with respect to the "principal" component, the next column has the scores with respect to the "second principal" component, and so on [54]. Based upon these "scores", each image can be projected onto the plane with the largest variation in subject movement. Figure 25 is an illustration of resulting images after projecting their pixels onto a specific plane based upon "scores" obtained from the PCA. Once the images have been made planar with respect to the camera, the next step is to negate any variations in subject data that may be caused by the varying physique and location of each subject. As explained in the next section, by normalizing all segmented images, a consistent size

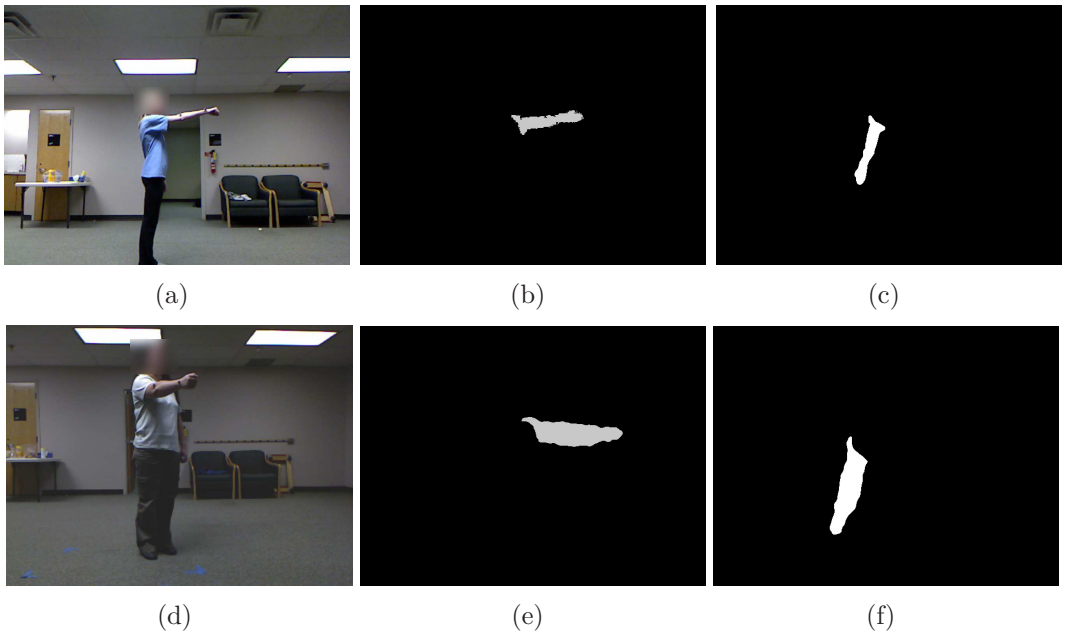


Figure 25: Illustration of PCA projected images: (a)original side image, (b)side image after Gaussian blur, (c)side image after PCA transformation, (d)original 45° image, (e)45° image after Gaussian blur, and (f)45° image after PCA transformation.

and location is ensured, thus proper comparisons are feasible.

2.7 Preprocessing for Real World Data Extraction

In order to accurately compare the human subject's dataset with a groundtruth dataset, each image needs to be normalized to a standard size and location within the image field of view (FOV). This process will ensure that the images of a human subject with a different arm-span than that of the previously stored images will be invariant in regards to size and location. Therefore, the Dynamic Time Warping process used to determine whether or not the human is performing the proper exercise (discussed in Section 2.9), will not be skewed due to these variations. Typically, image normalization is achieved by applying the image moments and affine transformations on each image in order to eliminate shearing and finally applying a scaling factor to ensure a standard size. This process in given in detail in the following sections and can be found in [36].

2.7.1 Image Moments and Affine Transforms

The details of the image normalization procedure were adopted directly from [36]. Let $I(x, y)$ denote a digital image of size M x N. Its geometric moments m_{pq} and central moments $\mu_{pq}, p, q = 0, 1, 2, ...$ are defined, respectively, as:

$$m_{pq} = \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} x^p y^q I(x, y) \tag{20}$$

and

$$\mu_{pq} = \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} (x - \overline{x})^p (y - \overline{y})^q I(x, y) \tag{21}$$

where

$$\overline{x} = \frac{m_{10}}{m_{00}}, \overline{y} = \frac{m_{01}}{m_{00}}. \tag{22}$$

An image $I^*(x, y)$ is said to be an affine transform of $I(x, y)$ if there is a matrix $\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$ and vector $d = \begin{pmatrix} d_1 \\ d_2 \end{pmatrix}$ such that $I^*(x, y) = I(x_a, y_a)$, where

$$\begin{pmatrix} x_a \\ y_a \end{pmatrix} = \mathbf{A} \cdot \begin{pmatrix} x \\ y \end{pmatrix} - \mathbf{d}. \tag{23}$$

Other examples of affine transforms include: 1) shearing in the x direction, which corresponds to $\mathbf{A} = \begin{pmatrix} 1 & \beta \\ 0 & 1 \end{pmatrix} \triangleq \mathbf{A}_x$; 2) shearing in the y direction, $\mathbf{A} = \begin{pmatrix} 1 & 0 \\ \gamma & 1 \end{pmatrix} \triangleq \mathbf{A}_x$; 3) scaling in both x and y directions, which corresponds to $\mathbf{A} = \begin{pmatrix} \alpha & 0 \\ 0 & \delta \end{pmatrix} \triangleq \mathbf{A}_s$. Where β, γ, α, and δ are arbitrarily defined shearing and scaling factors. Moreover, it is straightforward to show that any affine transform $\mathbf{A}$ can be decomposed as a composition of the aforementioned three transforms, e.g., $\mathbf{A} = \mathbf{A}_s \cdot \mathbf{A}_y \cdot \mathbf{A}_x$, provided that $a_{11} \neq 0$ and $det(\mathbf{A}) \neq 0$.

2.7.2 Image Normalization

The general concept of image normalization using moments is well-known in pattern recognition problems [94, 99, 100], where the idea is to extract image features that are invariant to affine transforms. In this application, a normalization procedure is applied to the image so that it meets a set of predefined moment criteria. The normalization procedure consists of the following steps for a given image $I(x, y)$.

1. Center the image $I(x, y)$; this is achieved by setting matrix A in Equation (23) to $\mathbf{A} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ and the vector $d = \begin{pmatrix} d_1 \\ d_2 \end{pmatrix}$ with

$$d_1 = \frac{m_{10}}{m_{00}}, d_2 = \frac{m_{01}}{m_{00}} \tag{24}$$

   where m_{10}, m_{01}, and m_{00} are the moments of $I(x, y)$ as defined in Equation (20). This step is aimed to achieve translation invariance. Let $I_1(x, y)$ denote the resulting centered image.

2. Apply a shearing transform to $I_1(x, y)$ in the x direction with matrix $\mathbf{A}_x = \begin{pmatrix} 1 & \beta \\ 0 & 1 \end{pmatrix}$ so that the resulting image, denoted by $I_2(x, y) \triangleq \mathbf{A}_x[I_1(x, y)]$ achieves $\mu_{30}^{(2)} = 0$, where the superscript is used to denote $I_2(x, y)$. This step is aimed to leave horizontal lines invariant, but maps the vertical lines into the new plane.

3. Apply a shearing transform to $I_2(x, y)$ in the y direction with matrix $\mathbf{A}_y = \begin{pmatrix} 1 & 0 \\ \gamma & 1 \end{pmatrix}$ so that the resulting image, denoted by $I_3(x, y) \triangleq \mathbf{A}_y[I_2(x, y)]$ achieves $\mu_{11}^{(3)} = 0$, where the superscript is used to denote $I_3(x, y)$. This step is aimed to leave vertical lines invariant, but maps the vertical lines into the new plane.

4. Apply a scaling to $I_3(x, y)$ in both the x and y directions with matrix $\mathbf{A}_s = \begin{pmatrix} \alpha & 0 \\ 0 & \delta \end{pmatrix}$ so that the resulting image, denoted by $I_4(x, y) \triangleq \mathbf{A}_s[I_3(x, y)]$, is

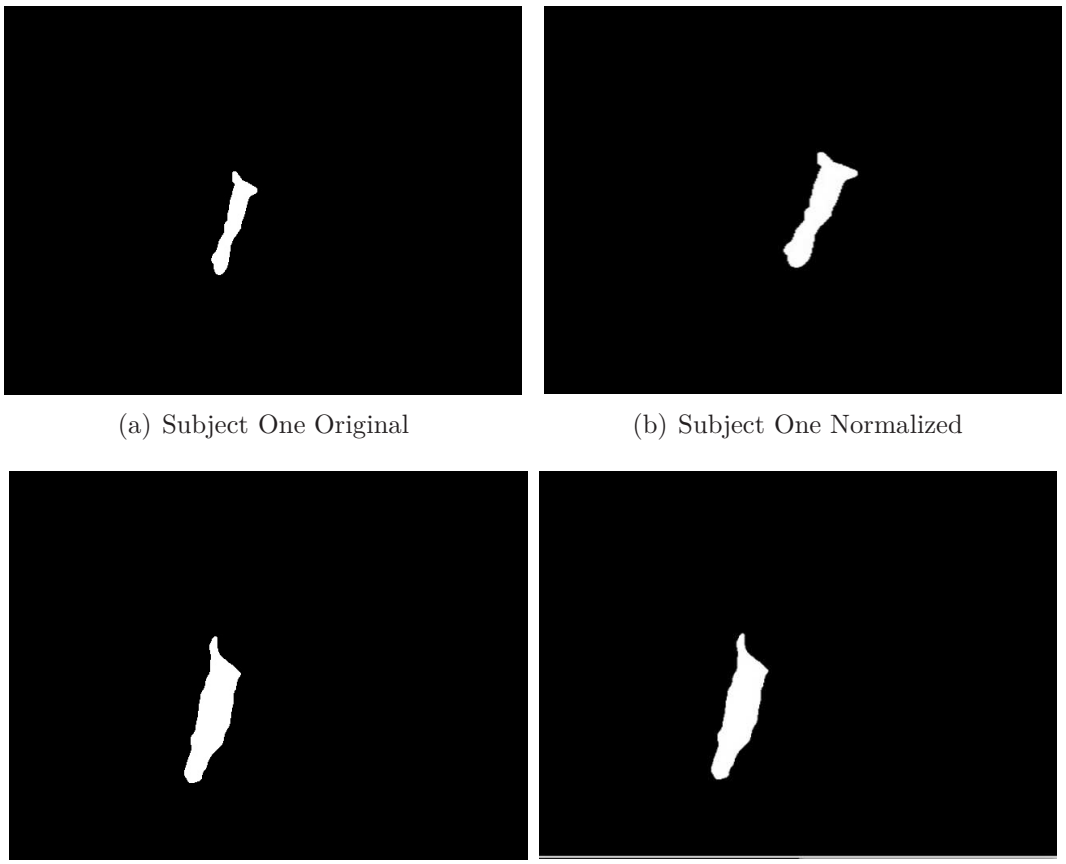
scaled to 1) a prescribed standard size and 2) $\mu_{50}^{(4)} > 0$ and $\mu_{05}^{(4)} > 0$, where the superscripts are used to denote $I_4(x, y)$.

The final image $I_4(x, y)$ is the normalized image, based upon the standard size extracted by the correlation dataset. The four steps in the normalization procedure are designed to eliminate each of these distortion components. More specifically, step 1) ensures translation invariance by preserving ratios of distances between points lying on a straight line by setting the center of the normalized image at the density center of the original image, steps 2) and 3) eliminate shearing in the x and y directions ensuring rotation invariance, and step 4) eliminates scaling distortion by forcing the normalized image to fit to a standard size [36].

The scale factor for each subject was created by taking the average pixel width and height of each subject's segmented movements and determining the factor of difference between the ground truth width and height averages obtained *a priori*. The pixel width and height of the segmented movements were determined by using a bounding rectangle based upon the top-most, bottom-most, left-most, and right-most pixels within an isolated segmentation. Also, calculating the average image centroid remains the same. This information can be used to scale and shift each image representing the subject to a standard size and location. Figures 26b and 26d illustrate the application of this process on the images from Figures 26a and 26c respectively.

2.8 Feature Vector

Once the subject's arm has been segmented and normalized for a specific frame, a feature vector is created for that frame by dividing it into an 8x8 grid, calculating the mean μ and standard deviation σ from each grid, and then constructing a 64 length feature vector. Various size grids could be used, however it is ideal to adhere to the law of diminishing returns by utilizing a feature vector that will allow the



(a) Subject One Original (b) Subject One Normalized

(c) Subject Two Original (d) Subject Two Normalized

Figure 26: Application of the image normalization process on data obtained from two separate subjects performing the same exercise. a) and b) are the original and normalized images of subject one, respectively, and c) and d) are the original and normalized images of subject two, respectively.

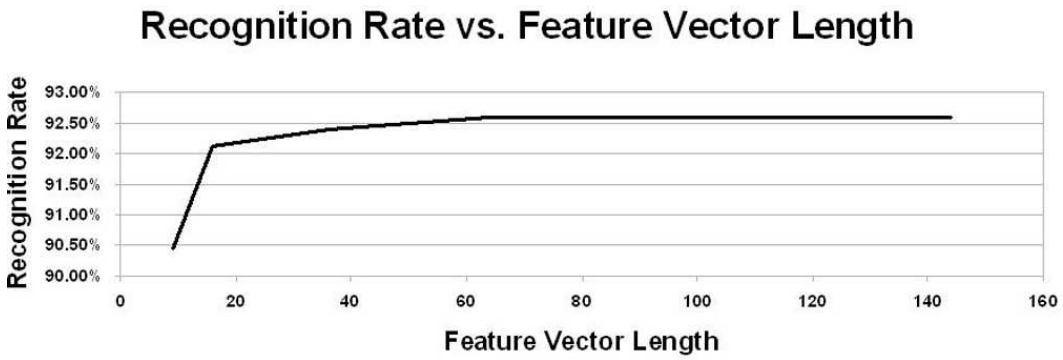


Figure 27: Graphical representation of feature vector selection.

Table 1: Resulting Processing Times and Recognition Rates for Various Feature Vector Lengths

Feature Vector Length	Processing Time (s)	Recognition Rate(%)
9	76.59	90.45
16	80.96	92.11
36	83.35	92.40
64	86.06	92.59
81	89.58	92.59
144	93.17	92.59

most features for image recognition without compromising a significant amount of processing time (i.e. minimize the classification error using the appropriate feature vector while maximizing the classification rate of the classifier). Figure 27 and Table 1 show the results of testing different size grids, which indicate that a 64 length feature vector is ideal.

Utilizing the feature vectors of the reference (or ground truth) images and the input (or currently segmented) images, a normalized Manhattan distance is calculated, for ease of use, between the two as shown in Equation (25):

$$d(\overrightarrow{x}, \overrightarrow{y}) = \sum_{i=1}^{n} |x_i - y_i| \tag{25}$$

where x and y are the feature vectors of the reference and input frames and d is the Manhattan distance. However, the normalized Manhattan distance only gives

information on a per frame basis. Once the subject's movements have been effectively represented, the next step involves aligning that movement with the groundtruth dataset.

2.9 Dynamic Time Warping

Dynamic Time Warping (DTW) is a process typically used in communication protocols for speech recognition [79]. The overall idea is to align the input speech data with data from a reference speech signal by repeating certain portions of a set of data that do not match the current position of the opposing set. For example, when someone wishes to place a phone call via voice dialing, he or she may simply say "Call, John Doe". The next time that the same individual wishes to call John Doe via voice dialing, his or her voice will undoubtedly have variations from the first time that voice dialing was initiated. For instance, the second time, the individual may say "Call, Jooohn Doe". Obviously, these two statements are not identical; however, the desire for the user to call John Doe has not changed. Therefore, DTW would be used to correctly match the vocal inflections made during the second speech (i.e. the input signal) with those of the previously stored first speech (i.e. the reference signal). Here, the three "o's" in the input signal would simply match with the single "o" in the reference signal, and John Doe would receive a phone call.

In this work, DTW is used to align image representations of robotic movements with those of the patient's movements. Since there will be variations in segment lengths and duration due to varying velocities between the two movements, this process becomes necessary for deciding whether or not two movements are similar. Figure 28 is a pictorial representation of the DTW process and is described as follows: 1) **Insertion**: The current reference frame does not match the current input frame, based on some threshold value γ, thus the current reference frame is repeated until a match is found during the linear search. 2) **Deletion**: The current input frame does

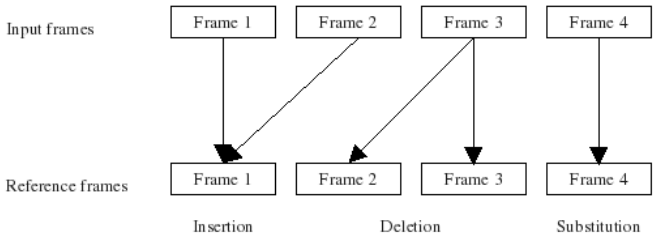


Figure 28: Pictorial representation of the basic operations of Dynamic Time Warping.

not match the current reference frame, thus the current input frame is repeated until a match is found. 3) **Substitution**: The current reference frame matches the current input frame, and the sequence progresses forward.

The Manhattan distances calculated from the feature vectors obtained during the segmentation process are stored in matrix form and used to determine the best possible choice between the operations for a specific frame comparison. In other words, the DTW matrix is populated by performing the following comparison:

$$dtw(m,n) = min \begin{cases} W_I d(m,n) + dtw(m-1,n) \\ W_D d(m,n) + dtw(m,n-1) \\ W_S d(m,n) + dtw(m-1,n-1) \end{cases} \tag{26}$$

where W_I, W_D, and W_S are weights for performing an "insertion", "deletion", or "substitution", d is the Manhattan distance between each frame, and m and n are the row and column positions in the matrices, respectively. The weights are chosen such that choosing to perform an "insertion" or "deletion" has more of a penalty than choosing to perform a "substitution" because the data has been somewhat skewed in choosing the former. It should be noted that the standard form of a DTW implementation assumes a known starting and ending point in the sequence; however, here, it is not assumed that the starting and ending points of the exercise are known. Thus, DTW is employed in order to determine a minimal exit point.

Utilizing this method, the least costly path (i.e. the optimal matching sequence) is calculated and a mapping between the input sequence to the reference sequence

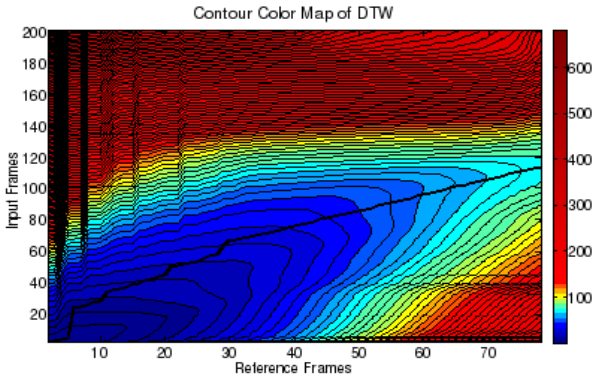


Figure 29: Contour color map of DTW illustrating the optimal path for mapping an input sequence to a reference sequence for one participant with a separate input sequence provided by the same participant. Given the blue contours represent points of lower cost, the black line shows the chosen path for mapping the two sequences.

is obtained thus minimizing the effects of varying velocities. In order to quantify this method for the purpose of this work, contour color maps were generated from patient exercise data. Figure 29 illustrates the results of mapping a reference sequence provided by the first participant with a separate input sequence provided by the same participant. As shown in the graph, the input sequence was significantly longer than the reference sequence; in essence, the participant performed the exercise at a faster velocity in the input sequence as opposed to the reference sequence. Given that darker contours represent points of lower cost, an optimal path would contain the maximum number of dark contours reaching the end of each sequence; the bold black line shows the chosen path for mapping the two sequences in this specific scenario.

As a comparison, Figure 30 illustrates the results of our DTW algorithm mapping a reference sequence provided by the first participant with a separate input sequence provided by the same participant performing two different exercises. There are two important points to note from the graph. First, although the DTW algorithm did find an "optimal" path for the two sequences, the path contains numerous light colored contours which are of high values. Second, only 50 frames of the 160 possible frames for the input image provided comparable feature vectors. In other words, this would

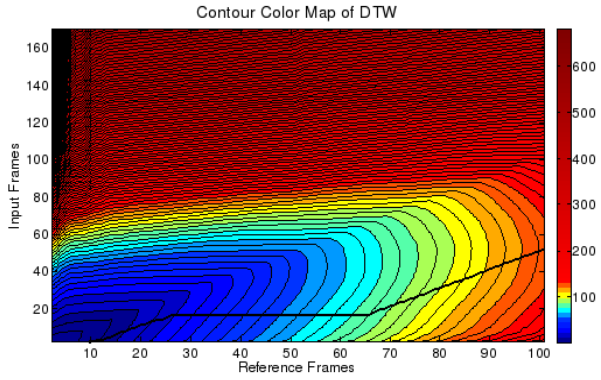


Figure 30: Contour color map of DTW illustrating the optimal path for mapping a reference sequence provided by the first participant with a separate input sequence provided by the same participant performing two different exercises.

not be considered a match for movements.

Once the total cost associated with the optimum path has been calculated, the next step is to determine whether or not the cost is below a certain threshold value. If the cost is higher then the threshold, then the subject repeats the exercise in hopes of ensuring proper movement. If the cost is below the threshold value, then the image processing continues to the next step of extracting contours representing the shape of the respective movements.

CHAPTER III

CALCULATING REHABILIATION PARAMETERS

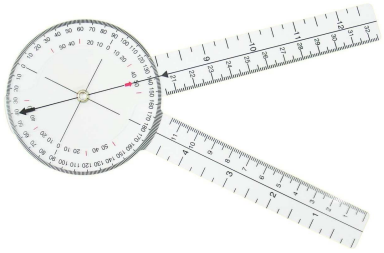


Figure 31: Illustration of a goniometer, a device used to measure the range of motion of a specific joint [69].

Once it has been determined that the subject has indeed performed the correct exercise, the next step involves calculating the kinematic metrics that pertain to the therapeutic assessment. By further applying image processing techniques to the previously segmented images of the arm's location throughout the exercise, mathematical properties can be used to extract prevalent information. Since this work focuses on non-contact, upper-arm rehabilitation for the glenohumeral joint (shoulder joint), prevalent information pertains to metrics that isolate and assess the shoulder's functionality. Traditionally, physical therapists have used (and still use) a measurement tool known as a goniometer, shown in Figure 31, to determine joint mobility; more specifically it determines the range of motion of a joint. When physical therapists want a more in depth analysis of the patient's mobility, or lack thereof, they typically use their own visual analysis based on experience or they turn to motion tracking via an expensive motion tracking system such as a Vicon system. In this chapter, it will be shown how computer vision techniques can be used to determine a patient's

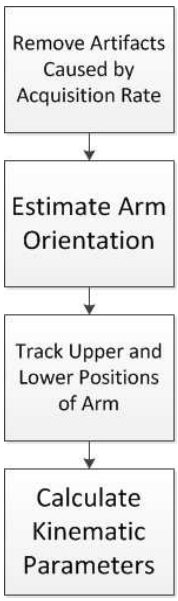


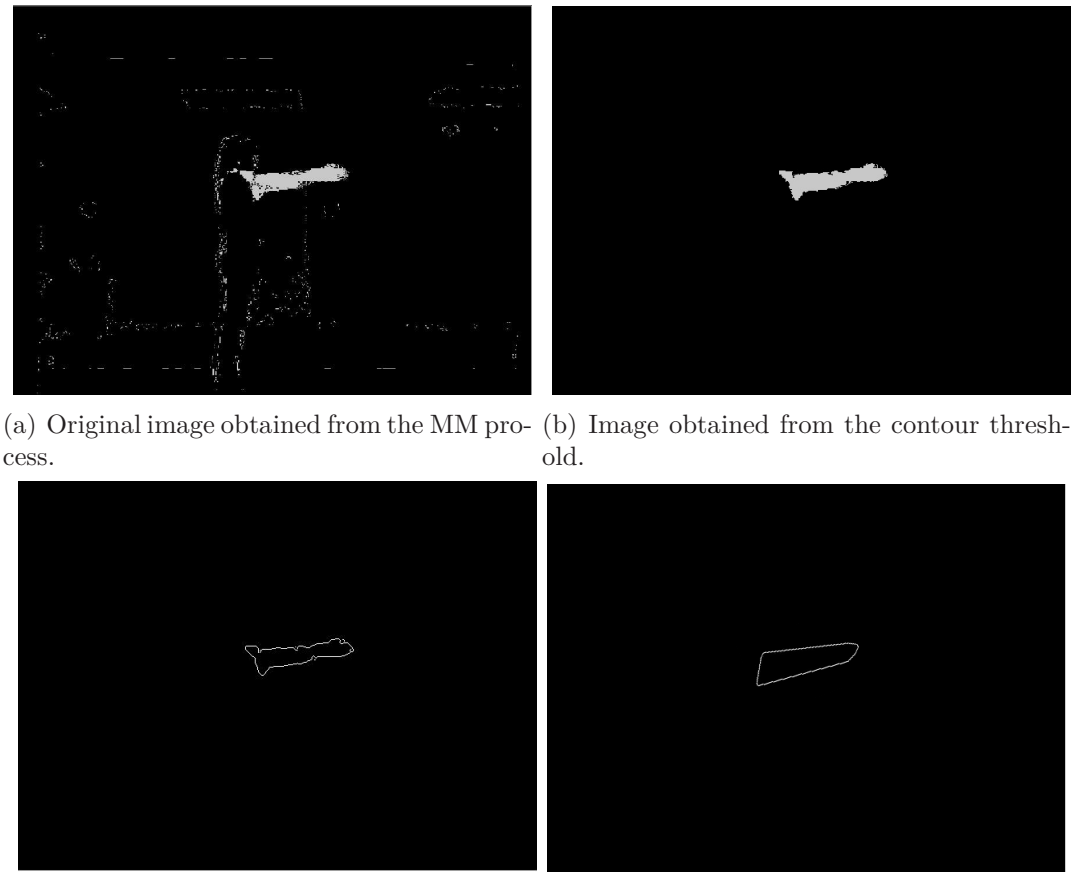
Figure 32: Chronological flowchart for calculating the kinematic metrics.

Active Range of Motion, Peak Angular Velocity, Movement Time, Total Displacement, Movement Units, and Normalized Jerk Score. These are all parameters that are used to assess a patient's ability to physically function in his or her activities of daily living. Figure 32 is a chronological flowchart of the steps taken for calculating the kinematic metrics.

3.1 Contour Extraction

The first step in this process is to utilize the images created by the contour threshold algorithm mentioned in Section 2.4. After the smaller contours have been removed, a canny edge detection algorithm [19] is utilized in order to extract the edges of the contour representing the upper-arm movement, see Figure 33c. Utilizing the edge detected shape, a proper representation of the sequence is then created. Again, due to camera inaccuracies there will undoubtedly be areas in the image where actual movement is not properly represented, even after the initial filtering and edge detection processes. Therefore, the convex hull of the edge detected image is determined.

The convex hull can be thought of as the boundary of a minimal convex set of



(a) Original image obtained from the MM process.

(b) Image obtained from the contour threshold.

(c) Image obtained from the Canny Edge Detection algorithm.

(d) Image obtained from the Convex Hull.

Figure 33: Example image processing sequence used to extract an ideal contour from the human's movements.

points containing a given non-empty finite set of points in a plane. The convex hull that is utilized is that of a simple polygon. By looking at three consecutive vertices of the polygon, during a recursive progression around the polygon, this algorithm simplifies to determining whether the resulting angle between the three vertices is concave or convex. If the resulting angle is concave, then the middle point is removed and the next (along the polygon) vertex is added to the triple to be tested. If the angle is convex, then each of the points in the triple is shifted by one vertex along the polygon [32]. Equation (27) shows the mathematical calculation for determining the convex hull of a polygon, where $H_{convex}(X)$ is the convex hull of set X. This implementation gives a somewhat ideal outline as seen in Figure 33d.

$$H_{convex}(X) = \left\{ \sum_{i=1}^{k} \alpha_i x_i | x_i \in X, \alpha_i \in \mathbb{R}, \alpha_i \geq 0, \sum_{i=1}^{k} \alpha_i = 1, k = 1, 2, ... \right\} \quad (27)$$

where α_i is non-negative and sums to one and x_i are a subset of the original points along the contour X. The equation basical states that the convex hull is the set of all convex combinations of finite subsets of points from the original contour X.

3.2 RANSAC

Now that a somewhat ideal outline has been obtained, determining the best method for finding the range of motion using only the image data is needed. Utilizing the major axis as a symmetrical dissection of the polygon and employing a Hough transform [37] on either the upper or lower region of the contour could enable a determination of the upper or lower line for the purpose of finding the angle between either line and the major axis. The Hough transform is a method used in computer vision to detect simple shapes, such as straight lines, by using the parameters of a line, $y = mx+b$ and representing the slope and intercept in parameter space (b, m). However, after much deliberation and testing, it was decided that since the Hough transform merely makes estimations of the best possible line to fit the upper or lower region and returns those estimations, a more accurate approach would be beneficial. Therefore, it was decided to use the **RAN**dom **SA**mple **C**onsensus (RANSAC) algorithm.

RANSAC determines the best possible line fit by iteratively selecting a random subset of the original input data and returns points from the original input data that are inliers. Given a set of data points U, there is an unknown number of data points that are consistent with a model with unknown parameters from parameter space Θ. These data points are inliers, and all others are outliers. The goal is to find model parameters θ^* from a parameter space Θ that maximizes a cost function J_S (θ, U, Δ). In the standard formulation, the cost function J_S is the size of the support of the model with parameters θ, i.e. how many data points from U are consistent with it.

Data points with error smaller than Δ are considered to be consistent or to support the model. An error function ρ (θ,x) representing the distance of a data point to a model is typically given. The threshold Δ is an input parameter to RANSAC [28,41].

The RANSAC algorithm carries out the maximization of J_S by repeatedly executing two steps: (i) a **hypothesis generation** step, where a hypothesis θ_k of the model parameters is computed from a subset of S_k selected from the input data points U at random and (ii) the **verification** step, where the quality of the hypothesized model parameters is calculated by utilizing a user predefined probability (confidence) 1 - η_o (typically set to 95%) to recover maximum of the cost function J_S [28, 41].

The algorithm can be described mathematically as follows: Let P be the probability that a sample of size m is randomly selected from a set U of N data points

$$P(I) = \frac{\begin{pmatrix} I \\ m \end{pmatrix}}{\begin{pmatrix} N \\ m \end{pmatrix}} = \prod_{j=0}^{m-1} \frac{I-j}{N-j} \leq \epsilon^m, \tag{28}$$

where ϵ is the fraction of inliers $\epsilon = I/N$. The number of inliers I is not known beforehand. Let χ_k^* be the largest support of a hypothesized model found up to the k-th sample inclusively, $I_k^* = |\chi_k^*|$. The sampling process is terminated when the likelihood of finding a better model (with larger support than I_k^*) falls under a threshold, i.e. when the probability η of missing a set of inliers χ^+ of size $|\chi^+| \leq I_k^*$ within k samples falls under a predefined threshold η_o,

$$\eta = (1 - P(I_k^*))^k \tag{29}$$

The number of samples that has to be drawn to satisfy $\eta \leq \eta_o$ is

$$k_{\eta_o}(I_k^*) = \frac{\ln(\eta_o)}{\ln(1 - P(I_k^*))}. \tag{30}$$

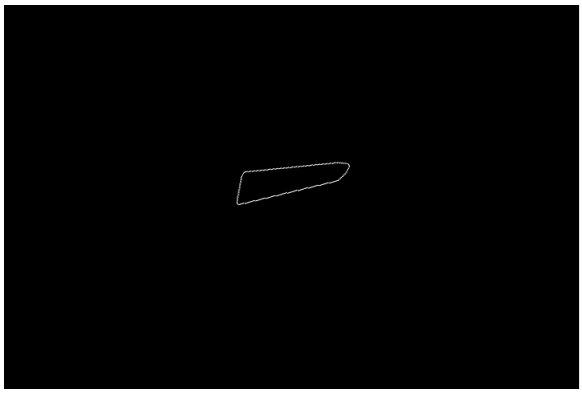


Figure 34: Illustration of the original convex hull image.

This method is a more accurate approach than the hough transform because it only returns points from the original input data (inliers) rather than creating its own values when predicting the line segment. Also, rather than using the major axis as one of the two lines used to find the range of motion, it was determined that a more accurate measure would be to perform RANSAC on the upper and lower regions (created by the major axis's dissection of the contour), thus creating a measure of the highest and lowest positions of the subject's arm. Figures 34, 35, and 36 illustrate this process.

3.3 Calculating Kinematic Metrics

3.3.1 Active Range Of Motion

Once the points that create the upper and lower lines are recognized, the slopes of each are used to calculate the angle between the two lines via simple geometry, shown in Equation (31):

$$\begin{aligned} m_1 &= \tfrac{y_2 - y_1}{x_2 - x_1} \\ m_2 &= \tfrac{y_4 - y_3}{x_4 - x_3} \\ \Theta &= \arctan(\tfrac{m_2 - m_1}{1 + m_2 m_1}) \\ AROM &= |\Theta \tfrac{180^\circ}{\pi}| \end{aligned} \tag{31}$$

where x and y are the coordinates of points on each line segment, m_1 and m_2 are the slopes of each line, and Θ and $AROM$ are the current angle in radians and degrees,

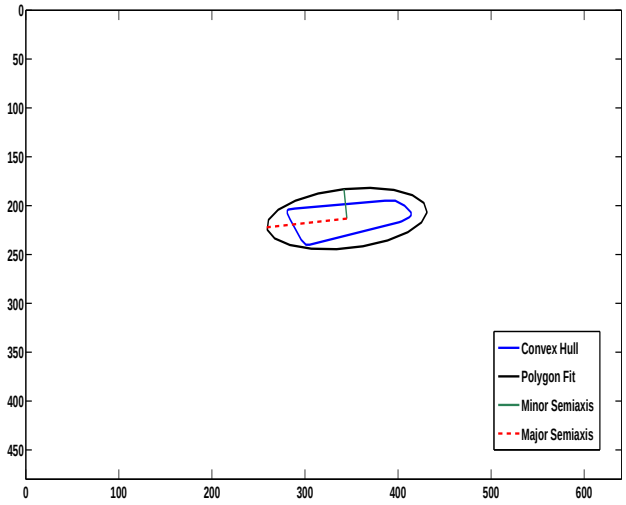


Figure 35: Illustration of the Major (dashed line) and Minor (dotted line) semiaxes located on a contour obtained from human upper-arm movement. The polygon boundary used to calculate the two semiaxes is shown with square markers.

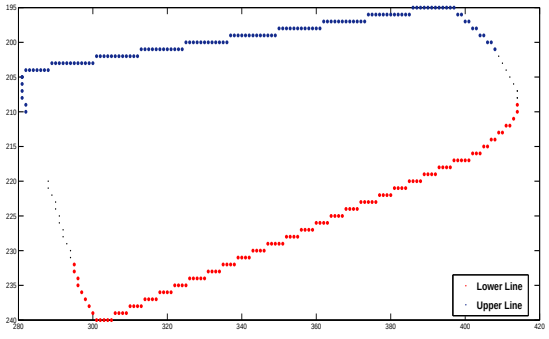


Figure 36: Lower line, determined by RANSAC, used to find the Active Range of Motion.

respectively. The maximum angle found over the length of the video sequence gives the Active Range Of Motion (AROM) of the patient's movements.

3.3.2 Peak Angular Velocity

Given that the frame rate of the camera used to capture the patient's movements was 15fps, calculating the angular velocity of the arm was straightforward. The initial lower line and subsequent upper lines that were recognized via RANSAC to determine the angular velocity as it relates to each frame was used. Meaning, the lower line found during the first RANSAC calculation over the convex hull in frame one of the video sequence was used as the initial position of the arm, while the current upper line changed as the subject moved his or her arm upward during the exercise. Using the standard equation for angular velocity, shown in Equation (32), the angular velocity, per frame, of the patient can be determined with the angle obtained from each pass of the AROM calculation and known frame rate.

$$\omega_i = \frac{d\theta}{dt} = \frac{\theta_i - \theta_{i-1}}{t_i - t_{i-1}}, \tag{32}$$

where $d\theta$ is the difference between the current and previous angles and dt is the frame rate. The maximum angular velocity would be considered the Peak Angular Velocity (PAV), shown in Equation (33).

$$\omega_{peak} = max(\omega_i) \; \forall \; i : i = 1...N \tag{33}$$

where N is the total number of angular velocities calculated for a set of images.

3.3.3 Reaching Kinematics

As mentioned before, reaching is typically analyzed by five kinematic characteristics: movement time (MT), total displacement (TD), peak velocity (PV), movement units (MU), and normalized jerk score (NJS) [26, 74]. MT is calculated by dividing the number of frames collected during movement by the known frame rate of the Kinect.

(a) Subject's starting position. (b) Subject's ending position.

Figure 37: Abduction and adduction movement demonstrated by participant.

The TD can be determined based on the world coordinate information already calculated as described in Section 2.5. To determine the TD, one only needs to calculate the distance between the beginning and ending position of the segmented arm during the motion. The logical approach is to calculate the Manhattan distance between the left-most non-zero pixel's beginning and ending position prior to applying the PCA.

PV can be calculated by dividing the TD by the time at each frame and taking the maximum value. However, calculating the peak angular velocity (PAV) after the PCA has been applied would allow better tracking and an easier calculation of the MU and NJS. The process described in Section 3.3.2 is used to calculate the PAV. MU are obtained from the acceleration and deceleration data during the motion; one acceleration and deceleration phase comprise one movement unit. The NJS is calculated as follows:

$$NJS = \sqrt{\frac{1}{2} \cdot \int j^2(t)dt \cdot \frac{d^5}{l^2}} \tag{34}$$

where j is the third time derivative of position data, d is the movement duration (MT), and l is the movement amplitude [1]. Thus, this approach gives several physical therapeutic metrics that can be easily read by physical therapists for the purpose of analyzing a patient's current status and overall progress.

(a) Subject's starting position. (b) Subject's ending position.

Figure 38: Lateral and medial movement demonstrated by participant.

3.4 Preliminary Experimentation

For the initial testing of the methodology of this research, two subjects were utilized, both male. The subjects were asked to perform a series of upper-arm exercises, which were captured via a simple webcam. The specific exercises involved adduction and abduction, shown in Figure 37, and lateral and medial movements, shown in Figure 38. The algorithmic approach was then used to process the images in order to obtain the AROM and angular velocity, which was then compared to the data obtained from separate subjects who performed visual observation techniques, and data captured via the Trimble 5606 Robotic Total Station. The specific task of the visual observers was to locate the top portion of each patient's arm in several sets of images. Using that information, in congruence with a constant location for the patient's beginning position, the AROM was determined and, by extension, the angular velocity as viewed by the visual observers.

The Trimble 5606 uses a "time-of-flight" measurement technique based on the pulse measurement principle; it measures the time for a very short transmitted pulse to travel to a targeted prism, held by the subject, and back, thus calculating the position of the subject's end-effector. As such, the data obtained from the robotic station represents ground truth. Tables 2, 3, 4, and 5 show the AROM comparison between the visual observers, algorithmic data, and Trimble 5606 for each subject per

Table 2: Measurements of AROM for Adduction/Abduction Exercise for Subject A.

Repetition	1	2	3
Visual Observer (°)	90.1	65.5	39.2
Algorithm (°)	86.4	56.5	40.1
Ground Truth (°)	90.3	56.5	31.1
Visual Observer Difference (°)	0.2	9.0	8.2
Algorithm Difference (°)	3.9	0.0	9.0
Visual Observer Error (%)	0.18	15.96	26.33
Algorithm Error (%)	4.32	0.05	29.07

Table 3: Measurements of AROM for Adduction/Abduction Exercise for Subject B.

Repetition	1	2	3
Visual Observer (°)	86.7	57.0	45.8
Algorithm (°)	88.1	54.9	45.6
Ground Truth (°)	90.0	65.0	42.1
Visual Observer Difference (°)	3.3	7.9	3.7
Algorithm Difference (°)	1.9	10.1	3.5
Visual Observer Error (%)	3.65	12.20	8.81
Algorithm Error (%)	2.14	15.53	8.43

exercise. Tables 6, 7, 8, and 9 show the PAV comparison between the aforementioned data sets.

3.5 Preliminary Results and Analysis

As shown in the tables, the AROM values calculated via our algorithm are closely related to the ground truth data (overall average visual observer error $< 11\%$ and overall average algorithmic error $< 9\%$). Figure 39 shows a comparison between the visual observers and algorithm, with respect to the ground truth data, of the averages across all subjects for the AROM data. For a patient with a limited range of motion, this algorithm could be used to identify the condition in real-time, given a known

Table 4: Measurements of AROM for Lateral/Medial Exercise Subject A.

Repetition	1	2	3
Visual Observer (°)	86.2	57.3	29.8
Algorithm (°)	80.5	55.5	28.0
Ground Truth (°)	90.0	67.2	26.4
Visual Observer Difference (°)	3.8	9.9	3.4
Algorithm Difference (°)	9.5	11.7	1.6
Visual Observer Error (%)	4.22	14.70	13.07
Algorithm Error (%)	10.57	17.43	6.25

Table 5: Measurements of AROM for Lateral/Medial Exercise Subject B.

Repetition	1	2	3
Visual Observer (°)	84.9	56.0	44.7
Algorithm (°)	84.5	62.8	38.4
Ground Truth (°)	88.4	63.2	40.5
Visual Observer Difference (°)	3.5	7.3	4.2
Algorithm Difference (°)	3.9	0.4	2.1
Visual Observer Error (%)	3.95	11.46	10.40
Algorithm Error (%)	4.42	0.67	5.11

Table 6: Measurements of PAV for Adduction/Abduction Exercise for Subject A.

Repetition	1	2	3
Visual Observer (ω)	82.1	62.2	57.0
Algorithm (ω)	79.6	57.7	67.4
Trimble 5606 (ω)	87.8	59.9	67.5
Visual Observer Difference (ω)	5.7	2.3	10.5
Algorithm Difference (ω)	8.2	2.2	0.1
Visual Observer Error (%)	6.51	3.77	15.53
Algorithm Error (%)	9.33	3.66	0.09

Table 7: Measurements of PAV for Adduction/Abduction Exercise for Subject B.

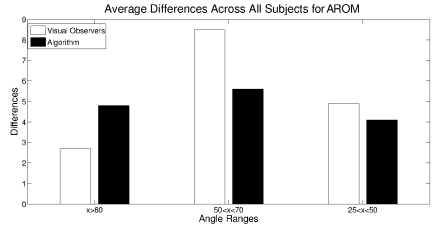
Repetition	1	2	3
Visual Observer (ω)	81.9	29.1	26.0
Algorithm (ω)	71.3	22.8	52.3
Trimble 5606 (ω)	62.4	23.9	54.6
Visual Observer Difference (ω)	19.5	5.2	28.6
Algorithm Difference (ω)	8.9	1.1	2.3
Visual Observer Error (%)	31.24	21.72	52.36
Algorithm Error (%)	14.32	4.41	4.13

Table 8: Measurements of PAV for Lateral/Medial Exercise Subject A.

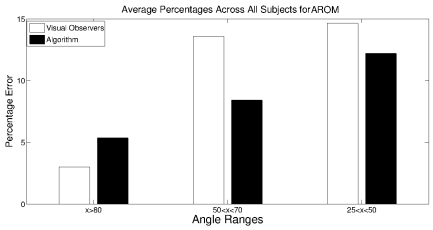
Repetition	1	2	3
Visual Observer (ω)	50.8	40.9	47.1
Algorithm (ω)	48.0	39.3	65.4
Trimble 5606 (ω)	51.3	43.8	65.9
Visual Observer Difference (ω)	0.5	2.9	18.8
Algorithm Difference (ω)	3.3	4.6	0.5
Visual Observer Error (%)	1.00	6.60	28.53
Algorithm Error (%)	6.48	10.42	0.74

Table 9: Measurements of PAV for Lateral/Medial Exercise Subject B.

Repetition	1	2	3
Visual Observer (ω)	39.7	37.3	30.8
Algorithm (ω)	34.1	40.2	27.4
Trimble 5606 (ω)	36.5	34.5	31.1
Visual Observer Difference (ω)	3.2	2.8	0.3
Algorithm Difference (ω)	2.4	5.7	3.7
Visual Observer Error (%)	8.83	8.05	1.09
Algorithm Error (%)	6.51	16.62	11.97

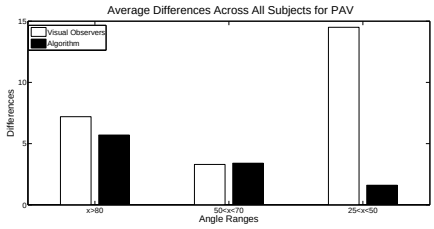


(a) Average Differences.

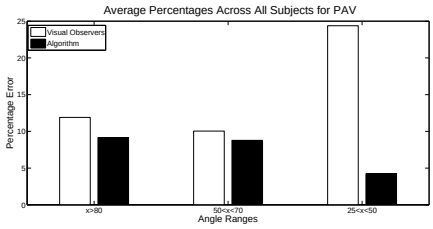


(b) Average Percentage Errors.

Figure 39: Comparison between the visual observers and algorithm, with respect to the ground truth data, of the averages across all subjects for the AROM data.



(a) Average Differences.



(b) Average Percentage Errors.

Figure 40: Comparison between the visual observers and algorithm, with respect to the ground truth data, of the averages across all subjects for the PAV data.

standard AROM. This will allow the system to monitor patient progress between sessions.

The PAV calculated via this approach is related to the trend of those calculated via ground truth data (average visual observer error $< 16\%$ and average algorithmic error $< 8\%$). Figure 40 shows a comparison between the visual observers and algorithm, with respect to the ground truth data, of the averages across all subjects for the AROM data. It should be noted that because the Trimble 5606 is a real-time tracking system, and human motion is not ideal, instances where the patient is not moving may not be conveyed with zero velocity in the ground truth data; thus values that are approximately zero are categorized as non-movement.

Of particular importance are the comments made by the visual observers when attempting to determine the range of motion of the patients. Both users noted that it was quite difficult to determine the exact location of Subject B's arm due to the fact that he was wearing a large blue sweater. One advantage that the algorithmic

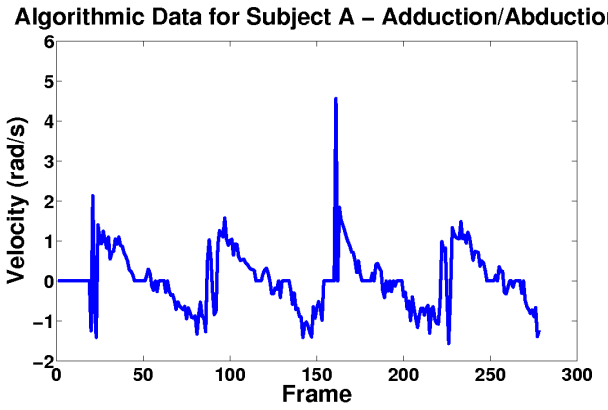


Figure 41: Plot of the raw angular velocity data of Subject A.

approach presented here has is that regardless of the clothing that the patient is wearing, the overall movement can still be tracked and reasonable data can still be obtained.

3.6 Gaussian Processes

Although the PAV is calculated via the previously described method, sparse data due to frame rate causes a plot of the raw angular velocity to appear noisy, see Figure 41. This measure can have negative affects when calculating the NJS which is dependent upon the derivatives of previously calculated position and angular velocity data. Thus, a process to alleviate this was warranted.

A Gaussian process (GP) is a collection of random variables, any finite number of which have consistent joint Gaussian distributions [88]. A GP is defined by a mean function, $\mu(\mathbf{x})$, which describes the mean output value of all possible sample functions evaluated at the input, $\mathbf{x}$, and a covariance function, $k(f(\mathbf{x}_i), f(\mathbf{x}_j))$, which specifies the correlation between pairs of random values. Although many covariance functions are possible, which allows prior knowledge of the function's behavior to be encoded in the GP framework, the typical choice is the squared exponential covariance function (also known as the Radial Basis Function (RBF)) shown in Equation (35).

$$k\left(f(\mathbf{x}_i), f(\mathbf{x}_j)\right) = \alpha \exp\left(-\frac{1}{2}\left(\mathbf{x}_i - \mathbf{x}_j\right)^T \Gamma \left(\mathbf{x}_i - \mathbf{x}_j\right)\right) \tag{35}$$

where Γ is a diagonal matrix of elements $\frac{1}{\gamma_1}, \ldots, \frac{1}{\gamma_N}$, and α is a scaling factor. In the 1D case, Equation (35) simplifies to:

$$k\left(f(x_i), f(x_j)\right) = \alpha \exp\left(-\frac{1}{2}\frac{\left(x_i - x_j\right)^2}{\sigma^2}\right) \tag{36}$$

where σ^2 is a scalar value equal to $\frac{1}{\gamma}$. The covariance between the outputs is written as a function of the inputs. This particular covariance function is derived from a Gaussian kernel, exhibits rotation and translation invariance to the inputs, and is infinitely differentiable or infinitely smooth. The exact spatial behavior of the covariance function can be tuned with a function-dependent set of parameters, known as hyperparameters in GP literature [88]. More specficially, the variables in the $N+1$ dimensional set $\alpha, \gamma_1, \ldots, \gamma_N$ are known as the hyperparameters for the squared exponential Gaussian process.

The specification of the covariance function implies a distribution over functions. This can be shown by drawing samples from the distribution of functions evaluated at any number of points. Despite the infinite dimensional nature of GPs, sampling is still feasible due to the marginalization property. If a GP is defined over a set, S, by $\mathcal{GP}(\mu, \Sigma)$, then the GP is also defined over any subset of S by the relevant submatrices of μ and Σ, as shown in Equation (37). Thus, as long as the number of points at which $f(\mathbf{x})$ is to be evaluated is finite, then sampling from the GP is also finite. Equivalently, any finite set of variables from a GP have a jointly Gaussian distribution [88]. An example of several functions drawn from a GP prior with zero mean and squared exponential covariance function is shown in Figure 42(a).

$$p(y_i, y_j) \sim \mathcal{N}\left(\begin{bmatrix}\mu_i \\ \mu_j\end{bmatrix}, \begin{bmatrix}\Sigma_{i,i} & \Sigma_{i,j} \\ \Sigma_{j,i} & \Sigma_{j,j}\end{bmatrix}\right) \Rightarrow p(y_i) \sim \mathcal{N}\left(\mu_i, \Sigma_{i,i}\right) \tag{37}$$

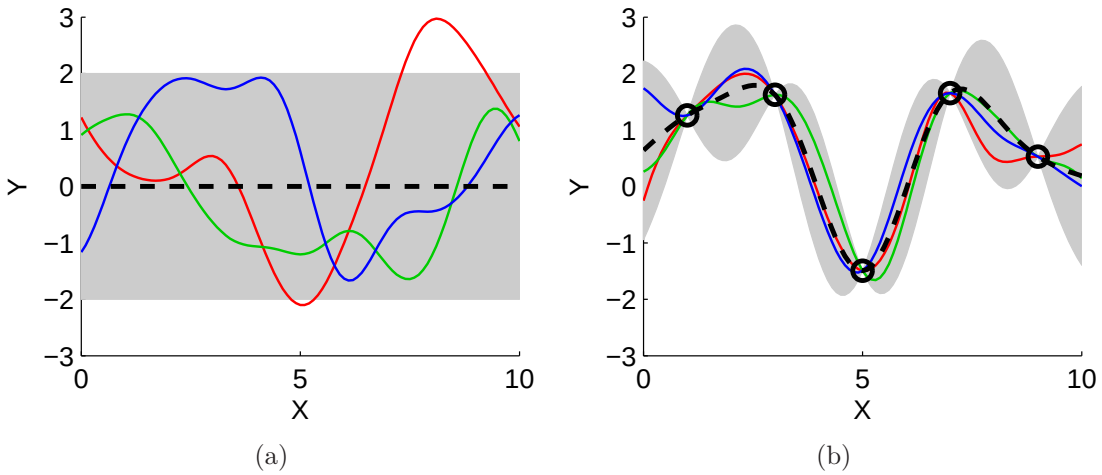


Figure 42: (a) An example of three functions sampled from a zero mean, unit variance Gaussian process, prior to applying any conditions. The mean is shown in black, and the 95% confidence area is shaded in gray. (b) An example of three functions sampled from a Gaussian process after conditioning on five measurement values. Again, the mean function is shown in black, and the 95% region has been shaded in gray.

A GP can also be conditioned on a set of known measurements [88]. The resulting GP posterior describes only the subset of sample functions that pass through the measurement points. This allows the GP to be used as a regression or interpolation technique, in which samples may be queried at an arbitrarily small resolution. However, unlike conventional regression or interpolation techniques, no data model (linear, quadratic, etc.) is required. For interpolation, a set of unknown output values, $Y^* = \left\{y_j^* | j = 1, \ldots, Q\right\}$, is desired, corresponding to a set of known inputs values, $X^* = \left\{\mathbf{x}_j^*\right\}$. The output values are to be conditioned on a set of known measurements, $Y = \{y_i | i = 1, \ldots, P\}$, corresponding to a second set of known input values, $X = \{\mathbf{x}_i\}$. The GP posterior mean and covariance satisfying these conditions are shown in Equation (39) and (40) (with a full derivation available in [110]). An illustration of the resulting posterior is shown in Figure 42(b).

$$p(Y^*|X, Y, X^*) \sim \mathcal{N}(\mu^*, \Sigma^*) \tag{38}$$

$$\mu^* = \mu_X + \Sigma_{Y,Y^*} \cdot \Sigma^{-1}_{Y^*,Y^*} \cdot (Y - \mu_X) \tag{39}$$

$$\Sigma^* = \Sigma_{Y,Y} - \Sigma_{Y,Y^*} \cdot \Sigma^{-1}_{Y^*,Y^*} \cdot \Sigma^T_{Y,Y^*} \tag{40}$$

where μ_S is a vector of values produced by evaluating the mean function, $\mu(\cdot)$, over the set, S, and Σ_{S_1,S_2} is a covariance matrix constructed by evaluating the covariance function, $k(\cdot,\cdot)$, with each pair-wise combination of values from sets S_1 and S_2.

By using the GP framework, derivative information can also be incorporated either as query points to be returned or as measurements to be used as conditions [101]. An output value, ω_j^m, is defined to be the partial derivative of the output function, $f(\cdot)$, with respect to the m^{th} dimension of the input, evaluated at the input value, $\mathbf{x}_j$, as shown in Equation (41).

$$\omega_j^m = \frac{\partial f(\mathbf{x})}{\partial x^m}|_{\mathbf{x}=\mathbf{x}_j} \tag{41}$$

The joint probability of the mixed vector containing y_i and ω_j^m involves evaluating the mean function, $\mu(\cdot)$, and covariance function, $k(\cdot,\cdot)$. If the mixed vector is used to condition the GP, then the mean values can be extracted directly from the measured values. However, if the mixed vector contains query points, then the mean values need to be derived from Equation (38). The covariance function, on the other hand, must be evaluated for all possible pairs contained inside the mixed vector for both cases. Equations (42) and (43) show how first derivative information may be incorporated into the existing problem structure, assuming the covariance function is differentiable [101].

$$k(\omega_i^m, y_j) = \frac{\partial}{\partial x^m} k(f(\mathbf{x}_i), f(\mathbf{x}_j)) \tag{42}$$

$$k\left(\omega_i^m, \omega_j^n\right) = \frac{\partial^2}{\partial x^m \partial x^n} k(f(\mathbf{x}_i), f(\mathbf{x}_j)) \tag{43}$$

Specifically, for the squared exponential covariance function in Equation (35), the possible modified covariance functions are listed in Equations (44) - (46).

$$
\begin{aligned}
k\left(\omega_i^m, y_j\right) &= -\frac{1}{\gamma_i}\left(x_i^m - x_j^m\right)\alpha\exp\left(-\frac{1}{2}\left(\mathbf{x_i}-\mathbf{x_j}\right)^T\Gamma\left(\mathbf{x_i}-\mathbf{x_j}\right)\right) && (44)\\
k\left(y_i, \omega_j^m\right) &= \frac{1}{\gamma_i}\left(x_i^m - x_j^m\right)\alpha\exp\left(-\frac{1}{2}\left(\mathbf{x_i}-\mathbf{x_j}\right)^T\Gamma\left(\mathbf{x_i}-\mathbf{x_j}\right)\right) && (45)\\
k\left(\omega_i^m, \omega_j^n\right) &= -\frac{1}{\gamma_i}\frac{1}{\gamma_j}\left(x_i^m - x_j^m\right)\left(x_i^n - x_j^n\right)\alpha\exp\left(-\frac{1}{2}\left(\mathbf{x_i}-\mathbf{x_j}\right)^T\Gamma\left(\mathbf{x_i}-\mathbf{x_j}\right)\right)\\
&\quad +\delta_{m,n}\frac{1}{\gamma_i}\alpha\exp\left(-\frac{1}{2}\left(\mathbf{x_i}-\mathbf{x_j}\right)^T\Gamma\left(\mathbf{x_i}-\mathbf{x_j}\right)\right) && (46)
\end{aligned}
$$

where $\delta_{m,n}$ is the Kronecker delta function [88], which is one when $m = n$ and zero otherwise. For the 1D case, Equations (44) - (46) would simplify to:

$$
\begin{aligned}
k\left(\omega_i, y_j\right) &= -\frac{1}{\sigma^2}\left(x_i - x_j\right)\alpha\exp\left(-\frac{1}{2}\frac{\left(x_i-x_j\right)^2}{\sigma^2}\right) && (47)\\
k\left(y_i, \omega_j\right) &= \frac{1}{\sigma^2}\left(x_i - x_j\right)\alpha\exp\left(-\frac{1}{2}\frac{\left(x_i-x_j\right)^2}{\sigma^2}\right) && (48)\\
k\left(\omega_i, \omega_j\right) &= -\frac{1}{\sigma^2}\frac{1}{\sigma^2}\left(x_i - x_j\right)\left(x_i - x_j\right)\alpha\exp\left(-\frac{1}{2}\frac{\left(x_i-x_j\right)^2}{\sigma^2}\right)\\
&\quad +\frac{1}{\sigma^2}\alpha exp\left(-\frac{1}{2}\frac{\left(x_i-x_j\right)^2}{\sigma^2}\right) && (49)
\end{aligned}
$$

A 1D example is shown in Figure 43. In this example, five evenly spaced position constraints were selected at random, along with four derivative constraints at a different spacing. The x-position of each measurement are stacked into the column vector, X, of length 9. The measurement values are similarly stacked into the column vector, Y. Y consists of both direct measurements, y_i, and derivative measurements, ω_i^1. Interpolated values for the function and the first derivative are requested at a resolution of 0.1 over the range $[0, 10]$. The 101 x-positions for the function output are stacked with the 101 x-positions for the derivative output to form the vector, X^*. The mean and covariance for the GP posterior are calculated using Equations (39) and

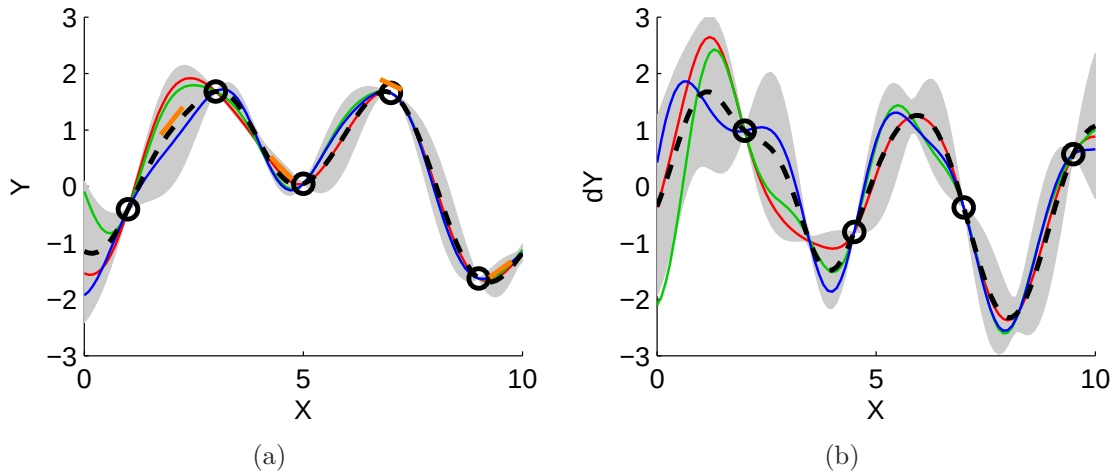
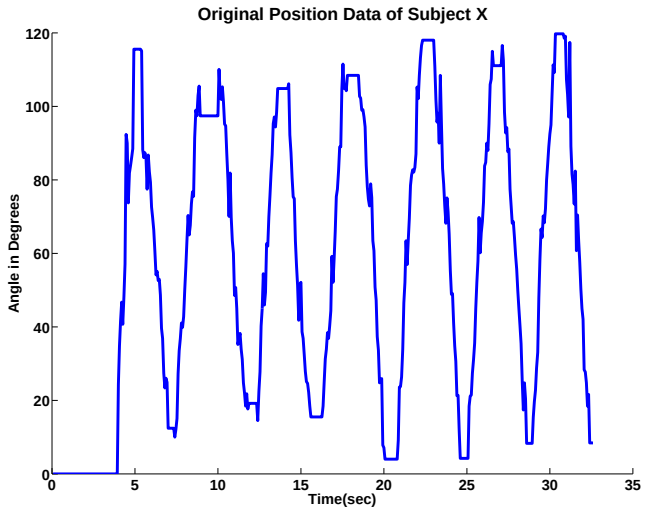


Figure 43: An example of three functions sampled from a Gaussian process after conditioning on five measurement values and four derivative values. The mean function is shown in black, and the 95% region has been shaded in gray. Both the function output (a) and the function derivative (b) are shown. Derivative constraints are indicated in the position plot as short, orange lines aligned in the direction of the derivative.

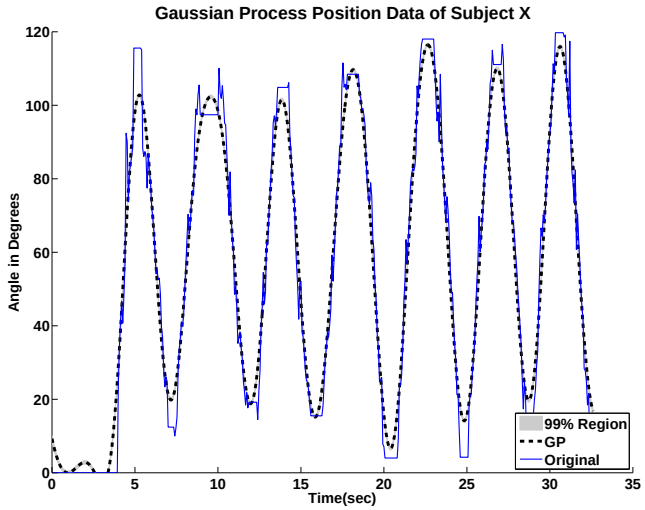
(40), where the individual entries in the covariance matrices $\Sigma_{Y,Y}$, Σ_{Y,Y^*}, and Σ_{Y^*,Y^*} are calculated using Equations (36) and (47) - (49). Functions can then be drawn from the GP by sampling from the 202-dimension multivariate Gaussian variable, $\mathcal{N}(\mu^*, \Sigma^*)$ defined in Equations (39) - (40) [115].

In the instance presented in this work, regarding the PAV data mentioned before, the GP can be used to calculate the derivative of the position data (AROM) in order to determine the PAV data of each subject. Furthermore, since the GP allows for multiple derivatives, the acceleration and jerk (third time derivative) can also be calculated. This process enables simple determinations of other important kinematic metrics, namely the MU and NJS. An example of the GP derived position data, velocity data, and acceleration data for the same subject's are shown in Figures 44 - 46.

Now that the algorithmic process has been solidfied, the next goal is to apply this method to various scenarios and determine its benefits within the field of robotic physical therapy. Comparing this approach with other novel methods and determining

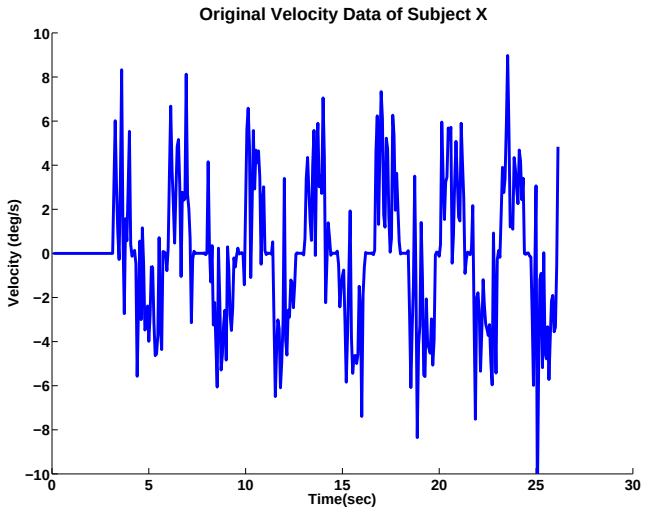


(a)

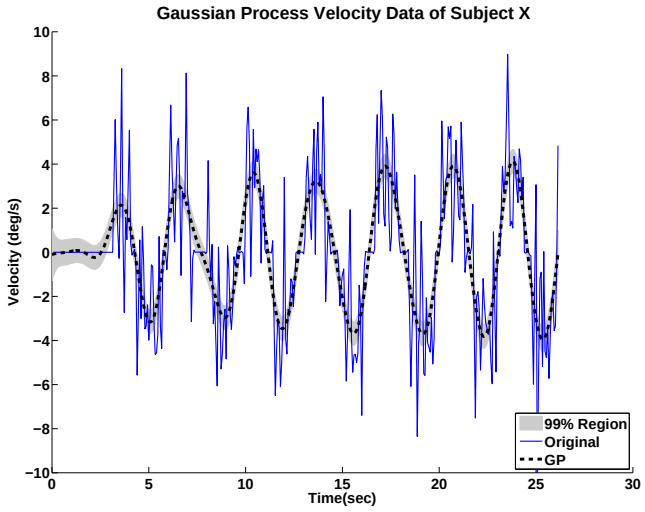


(b)

Figure 44: Results of the applying the GP to a subject's position data. The illustration shows the original position data (a) and the overlayed GP position data (b).

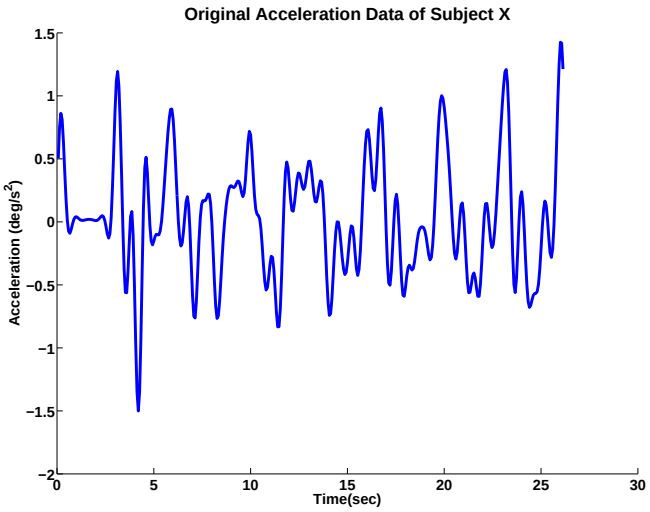


(a)

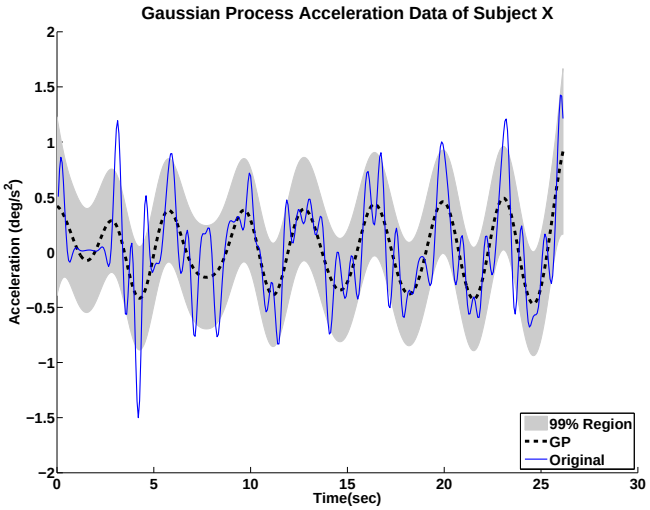


(b)

Figure 45: Results of the applying the GP to a subject's position data. The illustration shows the original velocity data (a) and the overlayed GP velocity data (b).



(a)



(b)

Figure 46: Results of the applying the GP to a subject's position data. The illustration shows the original acceleration data (a) and the overlayed GP acceleration data (b).

its potential applications is imperative to ensuring its employment in the field of robotic rehabilitation therapy.

CHAPTER IV

COMPARING CONTACT VERSUS NON-CONTACT REHABILITATION ASSESSMENT METHODS

As stated in the Introduction of this work, understanding and adapting to the needs of a specific patient during rehabilitation therapy requires continued research. While some patient's may suffer from complete paralysis of the upper-limb due to stroke, others may suffer from a less severe disability such as slight involuntary motions or tics. In order to cope with the range of child-rehabilitation objectives, a contact based approach of extracting upper-arm movement characteristics is compared with that of the non-contact, vision based approached described earlier. The contact method utilizes electromyography (EMG) sensors coupled with an HMM algorithm [75] to perform the same tasks as the non-contact method. Electromyography is a technique for evaluating and recording the electrical activity produced by skeletal muscles when the muscles cells are electrically or neurologically activated.

Comparisons of the two approaches show that the methodologies employed will depend on the capabilities of the user. For those with limited upper-limb movement control, the EMG system is ideal for enhancing muscle control associated with movement and would be valuable in a physical therapeutic facility whereas the vision-based system provides an low-cost solution to accurately assess simple movements, and therefore would be valuable for in-home use. This portion of the research was conducted in conjuction with Dr. Edward Brown at the Rochester Institute of Technology.

4.1 Contact Method for Upper-Limb Movement Extraction

The Hidden Markov Model (HMM) is a statistical process that models a system using a finite number of states. In the HMM process, these states are hidden from the user - the only data that is viewable is the sequence of observations. As such, the ultimate goal for the HMM is to predict the hidden sequence of states by only evaluating the set of observations. Since HMMs focus on stochastic processes, the technique could be used to predict upper extremity motions based on any associated EMG activity. The EMG signal represents the electrical activity that occurs when skeletal muscles expand and contract. EMG signals have been shown to be essential in the myoelectric control of artificial/prosthetic limbs [3], [55]. EMG signals provide pertinent information on the desired intent of muscle exertion as well as the types of motions the arm is commanded to perform. By observing a sequence of EMG signals extracted from muscles involved during upper-arm reaching motions, the HMM process could predict the actual joint angles related to the movement being performed (e.g., the joint angles of the elbow and shoulder). In this work, the sequence of EMG signals will correspond to observations the HMM will use as an input while the joint angle positions will be the sequence of states the algorithm will output.

Other research related to this approach include Chan and Englehart [24] who implemented an HMM classifier in order to distinguish between six forearm isometric contractions. They proved that, even with a low complexity HMM application, the classification rate yielded 94.5% accuracy on average. Another application of HMMs using myoelectric signals is described by Chiang, Wang, and McKeown [27]. The aim of the study was to distinguish between people who suffered from stroke versus those who did not. They collected surface electromyographic (sEMG) data and configured an HMM so each state represented a muscle group. Stroke patients muscle activity patterns were identified by the fact that specific muscle sequences were recruited when performing the motion of reaching. Kwon, Kim, and Choi explored ways to

combine an HMM algorithm with a multilayer perceptron (MLP) in order to improve the discrimination of the sEMG signals [61]. In their work, the upper-arm motions selected for classification included: elbow flexion and extension, wrist pronation and supination, and wrist flexion and extension. Feature extraction using the wavelet transform gave the best recognition rate. This method was only tested offline and was not extended to a real-time application.

4.1.1 Acquiring the EMG data signal

To implement the HMM algorithm described in this work, a data acquisition system capable of simultaneously capturing a subject's EMG activity during upper-arm motions and the joint angle information associated with those motions needed to be incorporated. The Upper Extremity Motion Capture System, which was designed and developed in the Biomechatronic Learning Laboratory, was used for this purpose [75]. The system, presented in Figure 47, consists of a platform for stability and an upper-arm brace. The brace constrains the arm to three degrees of freedom and is made of aluminum, while most of the platform is made from steel. The construction materials were selected to minimize the weight of the brace on the subjects' arms and to insure that EMG data could be properly recorded without altering the signal from additional loads. The system is configurable for both the left and right arm. The subject being tested may stand or sit when data is being recorded. To insure comfort, the brace is padded with vinyl covered foam, and nylon straps are used to keep the arm from sliding in the brace.

Joint angle measurements are taken through the use of three rotary potentiometers are integrated in the brace: two at the shoulder joint, and one at the elbow joint. Each potentiometer has a voltage range of +/- 2 volts, corresponding to joint angles between 0° - 360°. The top shoulder joint is configured to collect data from 0° - 140° in the transverse plane, while the side shoulder joint collects data in the sagittal

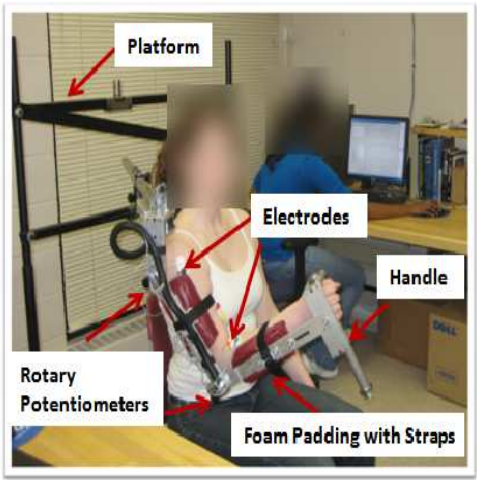


Figure 47: Upper Extremity Motion Capture System.

plane within the same range. The elbow joint can measure between 17° - 180° in the sagittal plane. EMG data is recorded through the use of dual EMG electrodes with two snaps and two gel sites separated by 1 cm. Both EMG and angle measurements are sent to the BioRadio^{TM}, a lightweight, commercial bioinstrumentation system with 12 configurable channels from Cleveland Medical Devices [29].

Since EMG data tends to be large and usually contains redundant information, the data is first transformed into a representative set of features to reduce the amount of data under analysis without compromising the information it contains. Since EMG signals are continuous, the data also needs to be quantized in order to be processed as a discrete signal. Hence, feature extraction needs to be performed on the raw EMG data, followed by a vector quantization before it is used as input to the HMM process.

4.1.2 Feature Extraction and Vector Quantization

Autoregressive (AR) modeling [118] is a technique that represents the present sample as a linear combination of past samples, plus an error term. Each past sample is multiplied by a coefficient before they are summed up. The coefficients form the

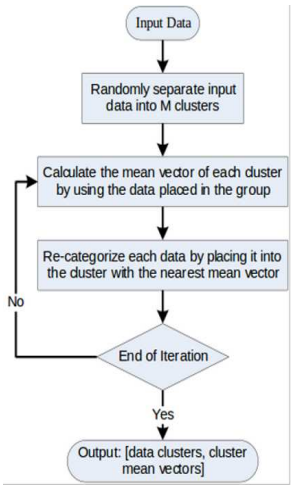


Figure 48: Flowchart of K-means Algorithm

autoregressive parameters. Equation (50) illustrates this process.

$$X_t = \sum_{i=1}^{M} \varphi_i X_{t-i} + \epsilon_t \tag{50}$$

where X_{t-i} are past inputs, φ_i is the autoregressive parameter, ϵ_t is the error term, X_t is the predicted output, and M is the total number of past inputs. In this work, the Burg algorithm [8,66], which is a widely used technique for extracting AR coefficients by estimating the reflection coefficients at successive orders, is applied. Once the AR coefficients are determined, the EMG signals to predict patient muscular activity via the HMM model is used.

Vector quantization involves mapping the input data space onto a finite set of observations in order to reduce redundant information while preserving important ones. Here, a K-means clustering process is employed in which each data point is assigned to the cluster whose center is the nearest. The center of a cluster is determined by averaging all the data points in the cluster. A flow chart of this algorithm is presented in Figure 48.

4.1.3 Hidden Markov Model

A Hidden Markov Model (HMM) has two primary elements, a transition matrix, A, and an emission matrix, B. Each element in the transition matrix, a_{ij}, signifies the probability of the transition from state i to state j as displayed in Equation (51) below:

$$a_{ij} = P(S_j|S_i) \tag{51}$$

For the emission matrix, each element, b_{ij}, corresponds to the probability of an observation, O_j, occurring given a state, S_i:

$$b_{ij} = P(O_j|S_i) \tag{52}$$

In summary, HMMs need to have a specification on the number of states, M, the number of observations, N, the transition matrix A, the emission matrix B, and the initial state distribution, π. These parameters are commonly represented by a compact notation, λ [24]:

$$\lambda = (\pi, A, B|M, N) \tag{53}$$

In order to develop all the necessary parameters of an HMM, three basic tasks need to be performed. The first task is to calculate the probability of the sequence of observations corresponding to the specific HMM parameters. This is known as evaluation. The second task involves the discovery of the sequence of states that generated the corresponding observations. This is referred to as the decoding stage, and the Viterbi algorithm is used to perform this task. Lastly, the third task is to adjust the model parameter, λ, in order to maximize the probability of a sequence of observations fitting a model. This stage is called the learning stage [87]. For this application, observations are associated with the quantized EMG signal and the state represents joint angle measurements. This approach is an extension of previous work shown in [75].

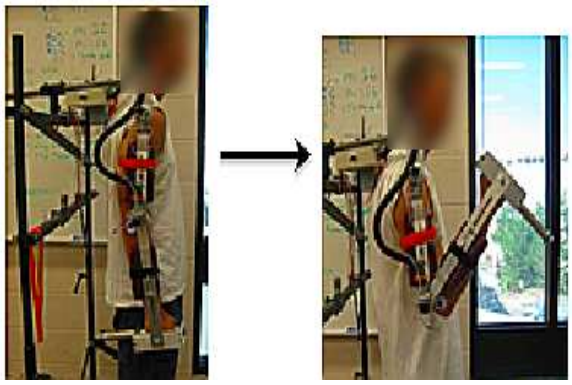


Figure 49: Elbow Flexion

Figure 50: Arm Extension

4.2 Experimental Protocol

For the contact-based protocol, the Upper Extremity Motion Capture system was used in conjunction with the CaptureLite^{TM} software from Cleveland Medical Devices to perform data collection. Data was acquired at a frequency of 960Hz with 12 bits data rate. The first two channels of the BioRadio^{TM} were configured to accept DC signals from the potentiometers representing the side shoulder and elbow joints. The next four channels were set to collect EMG signals from the deltoid, the biceps brachii, the triceps brachii, and the brachioradialis. For the non-contact based method, a simple webcam with a 15 frame per second (fps) frame rate was used to capture the motion of each subject and the aforementioned in Chapter 3 algorithm was used to process the data.

Figure 51: Shoulder Flexion

All data was collected from the right arm of five able subjects, three males and two females. Subjects were required to perform three different motions: shoulder flexion, elbow flexion, and arm extension. Shoulder flexion consisted of keeping the entire arm straight and rotating it about the shoulder joint in the sagittal plane. Elbow flexion corresponded to the flexion or the curling of the arm at the elbow joint. Arm extension is a combination of shoulder and elbow extension performed in the sagittal plane. The subject begins this motion starting approximately from the 90° elbow joint angle and then reaches forward in the sagittal plane. All three arm motions are presented in Figures 49, 50, and 51. For each motion, the subject was asked to perform three trials of thirty second long repetitive and continuous arm actions.

For the contact-based protocol, the first two trials were categorized as training data while the last trial was used for testing the HMM algorithm. Once data was retrieved, the HMM tracking algorithm, for the contact method, was developed in MATLAB®. The EMG signal is decomposed into segments of 100 data points as suggested by Oskoei [3]. Features were extracted using a fourth order autoregressive model, which collects the first four coefficients from each channel. Hence, each segment contained sixteen features. Vector quantization was performed using the K-means algorithm. The EMG training data was decomposed into 30 observations while the corresponding joint angle measurements were categorized into 20 states.

The training data also determined the K-means clustering mean vectors for the set of observations and states. The first trials of each motion were primarily used to estimate the parameters of the HMM. The second trials were then combined with the first trials in order to iteratively alter the HMM parameters using the Viterbi algorithm. This process modifies the model to make it a better representation of the data. Lastly, the EMG data from the testing group was quantized using the predetermined K-means clustering mean vectors. The output was then passed over to the HMM algorithm, which estimates the sequence of states that correspond to the input data.

For the non-contact based protocol, data collection was trivial. Given that the only data that is needed to assess subject movement is camera data, the subjects' movements as they were performing the required actions only needed to be recorded. In order to ensure that this approach was valid, the subjects were tracked using a Vicon Motion Capture System which incorporates infrared reflective markers for infrared camera tracking at 100 Hz. Once the camera data was collected, the non-contact algorithm was used to determine the subjects' joint angle positions and compared to the actual measurements collected from the Vicon Motion Capture System simultaneously (indicated by the word 'actual' in the figures). Markers for the Vicon Motion Capture System were placed on the shoulder, elbow, and wrist.

The results from each algorithmic approach (HMM and computer vision) were plotted against the actual joint angle measurements to evaluate its accuracy. Figures 52 and 53 show the elbow flexion exercise for a single subject as measured by the HMM and computer vision algorithms respectively. It is worth noting that the measured angles become flat at both peaks and valleys for each cycle because we retain the previous value during instances of motion that are lower than a certain threshold (i.e. the subject has basically paused before beginning the next half of the cycle).

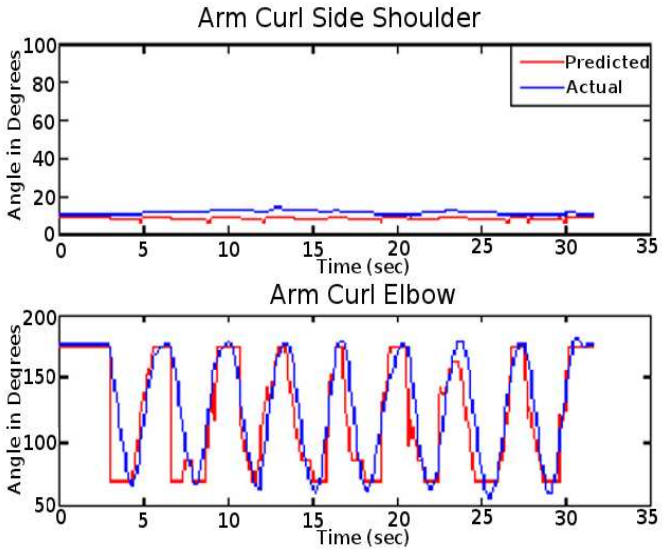


Figure 52: Motion Tracking Using HMM for Elbow Flexion

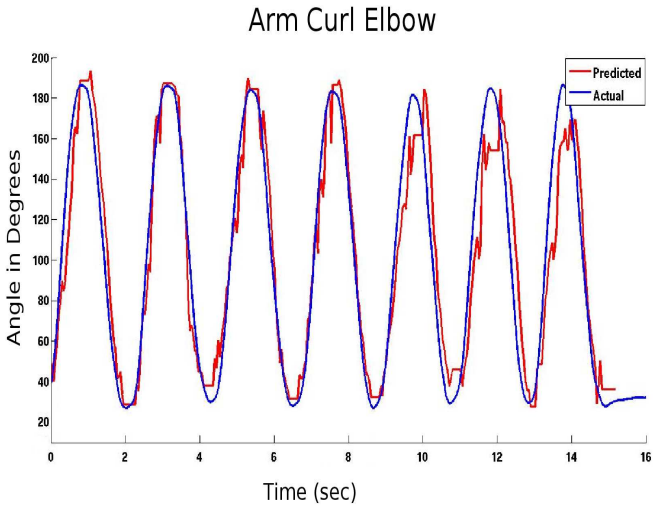
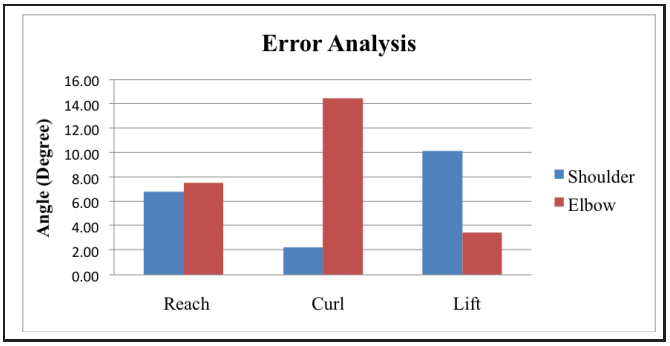
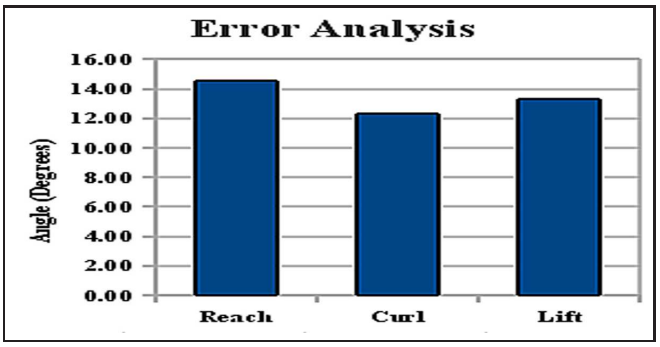


Figure 53: Motion Tracking Using Computer Vision for Elbow Flexion



(a) HMM Algorithm



(b) Computer Vision Algorithm

Figure 54: Error Analysis for Five Subjects.

To evaluate each algorithm, error analysis was performed by calculating the average error between the predicted and actual angle measurements. The mean error for the five subjects was then calculated in order to create a tabular representation of the error analysis. Graphs depicting the mean error for the HMM algorithm and the computer vision algorithm are presented in Figure 54. With overall mean errors of 14° and 13° for the HMM and computer vision algorithms respectively, both options are viable for calculating subject movements during robot-assisted rehabilitation. As a comparison, the work presented by Au and Kirsch [4], which follows the same procedure (as opposed to other works) as the one described here, yielded an error less that 20°. The largest error in both the HMM approach and the work by Au and Kirsch was found during arm curling/elbow flexion motions. Whereas, the largest error for the computer vision approach was found during the reaching motion.

4.3 Advantages and Disadvantages

Given that both methods perform reasonably well, it is prudent to weigh the pros and cons of each when determining which to employ in a rehabilitation scenario. The main issue in the development of the HMM algorithm was to determine the optimum observation and state numbers that the K-means algorithm should produce. Experimental testing revealed that the combination of 30 observations and 20 states yields the optimum result. Therefore, the HMM model requires one to have access to EMG sensors as well as allow time for some amount of learning/training.

Even though the computer vision model simply required a camera for data acquisition, one of the main obstacle was determining an appropriate value for the maximum component difference threshold and the cell threshold in the Multimodal Mean process. These values are important because the overall contour of the subject's movement depends on the shape that has been extracted via the Multimodal Mean.

The HMM approach is a more robust model for tracking upper extremity joint angles. Utilizing multiple EMG sensors, which give detailed information about the patient's muscular activity, gives the ability to track more complicated movements. For instance, in order to extract information from both the shoulder and elbow during a motion that requires both joints to move simultaneously using the computer vision approach, one may need to either create a model of the patient that would separately encapsulate line segments from both positions or consider each joint movement as separate entities. This is not possible in the current vision system. However, for simple motions, which either requires only one joint to move significantly or monitoring the movement as a whole, the computer vision approach allows for quick and reliable calculations with minimal cost.

Finally, it is also prudent to note the feasibility of patients being able to obtain each device for personal use. Obviously, the HMM approach requires a pricey motion capture system that consists of EMG sensors, a platform for stability, and an

upper-arm brace that would be challenging for in-home use. However, it is ideal for rehabilitative facilities requiring devices for children suffering from hemiplegia or hemiparesis. In contrast, the computer vision approach is a cost efficient method for in-home rehabilitation. With the purchase of a simple webcam and software, a capable patient could perform rehabilitative exercises at home and have the results sent to his or her physical therapist for analysis. Through this medium, physical therapists would be able to track the progress of patients throughout the recovery period both inside and outside of outpatient therapy. Still, the idea of being able to perform self-directed, in-home rehabilitation beckons the question as to whether or not patients properly comply to the directions of the physician once they are at home. In other words, even with the ability to monitor and track progress, it is still the responsibility of the patient to perform the required exercises during his or her own time. Perhaps there are methods to aid patients in the process of isolated, in-home therapy, which is the focus of discussion in the next chapter.

CHAPTER V

DETERMINING PROPER COMPLIANCE TO INSTRUCTIONAL UPPER-ARM EXERCISES

The inability to properly adhere to a physicians instructions could result in a delayed recovery or, potentially, a chronic issue. Therefore, proper compliance to instructional exercises is vital to the overall well-being of the patient. With regards to incorporating robotic assistance, specifically for non-contact interaction, determining proper compliance could ensure that the patient reaches full recovery in a timely fashion.

In Section 2.9, it was discussed how DTW could be used to determine whether or not a subject was performing the same general exercise that a robot was demonstrating based upon previously recorded and stored images. At its core, this process depends upon user-defined thresholding to determine if the two sets of data are similar. A more concrete analysis could be made based upon the kinematic metrics associated with the specific exercises. More specifically, comparing the kinematic data from a ground truth set of images to those of the patient may give more insight regarding compliance. The ground truth kinematics could be extracted from the images of an expert, say a physical therapist (PT), and be incorporated into a robotic platform via inverse kinematics for demonstrating those actions to the patient at a later time without the PT being present. This could ensure proper compliance even in self-directed, in-home exercises.

With this notion in mind, the question arises as to whether or not an embodied robotic platform is even necessary. One could simply mimic an instructional video while recording his or her movements and determine the kinematics within that scenario. The underlying question, one that has been discussed for years, with regards

to the field of robotics becomes, "Physical robots or virtual agents?" Which is better for improving back-and-forth interaction with humans, whether for education [5,107], rehabilitation [17, 49], or just gaming [78, 106]? With the advances in visualization and realism of virtual agents in 3D environments, some believe that much of what robots achieve with regards to human-agent interaction can also be achieved with an intelligent, less expensive, simulated version. Although robot enthusiasts would argue otherwise, by defending the practical and theoretical importance of physical embodiment in human-agent interaction, there are very limited studies that formalize the benefits in using physical agents versus virtual agents.

As such, the research presented in this chapter takes a novel approach of applying the aformentioned computer vision algorithms coupled with human-agent interaction to determine the significance of using a simulated agent versus an embodied agent during the mimicking of a physical exercise. The chapter begins with a brief background regarding other research focused on simulated versus embodied agents and the methods scientists have used to analyze their studies. Then, the experimental protocol for this particular study is outlined in Section 5.3, and the results of the study are analyzed in Section 5.4. Finally, Section 5.5 concludes the chapter with a discussion and future implications.

5.1 Background

The majority of the studies that address the issue of simulated versus embodied agents typically focus on comparing differences based on elements of human perception and engagement. To date, researchers have solely used a survey analysis approach to derive an answer to the simulated versus physical agent question. For example, Takeuchi et al. [104] wanted to determine whether or not there were any clear advantages to using a simulated agent versus an embodied agent when presenting information to a human audience. The authors used an on-screen (simulated) agent in comparison

to ASIMO (the embodied agent) to present a weather forecast using a multimodal presentation markup language. Using a post-session questionnaire, the authors found that ASIMO was rated higher in areas such as human likeness, goodness of the presentation, and the user's overall interest. However, the on-screen agent received higher scores for participant comprehension, focus on the presentation, and the agent's ability to accurately point to the objects.

Powers et al. [84] researched differences between simulated and embodied agents as those differences pertain to the disclosure of health or sensitive information. Using four different scenarios – Computer agent on a computer monitor, computer agent projected life-size on a screen, remote robot projected life-size on a screen, and collocated robot – in a between-subjects test, participants answered certain questions regarding general health habits. The researchers deduced that choosing between an embodied or simulated agent was very task specific. For tasks that involve a significant amount of information transmission but relatively little social rapport (e.g., information kiosks), disembodied agents should suffice. Likewise, for tasks that require users to reveal personal information, disembodied agents may be preferable. However, for tasks that are more relationship-oriented (e.g., a home companion), a collocated robot would seem to be better suited.

Finally, Lee et al. [62] questioned the significance of embodiment as it pertains to social agents and their tactile interaction. Using the touch sensors on the Sony Aibo [77, 85] in comparison to mouse clicks when interacting with a virtual version of the same platform, subjects were allowed to interact with their respective agent for 10 minutes. Again using a post test survey, the researchers determined that physical embodiment has an added value for people's social interaction with agents by increasing the social presence.

While each of the previously described bodies of work give insight on this topic with respect to social interaction, none incorporate a physical metric based approach

Table 10: Survey Questionnaire from Earlier Pilot Study

Q1	On a scale from 1 to 5 (5 being the highest level), how would you quantify your enjoyment level from the experience with: 1.) the simulated robot 2.) the actual robot?
Q2	On a scale from 1 to 10 (10 being the highest level), how difficult was it to recognize the actions of: 1.) the simulated robot 2.) the actual robot?
Q3	On a scale from 1 to 10 (10 being the highest level), how difficult was it for you to repeat the actions of: 1.) the simulated robot 2.) the actual robot?
Q4	Do you think the simulated (or actual) robot would be beneficial for rehabilitation therapy of the upper limb?

that would provide a more concrete analysis. More specifically, for therapeutic robotics, a metric based approach would provide a more concrete analysis regarding the physical attributes and effectiveness of the overall treatment. Therefore, the recent literature induces a question in the field of therapeutic robotics. Namely, "Does an embodied agent induce better human adherence to physical therapy exercise than a simulated agent regardless of user perception?" The intellectual merit of this question stems from the progression of using robotics to administer aid during physical rehabilitation [56, 59, 105].

5.2 Motivation

An earlier study was conducted in order to better understand human perception with respect to interacting with a simulated agent and an embodied agent in an exercise scenario. In this particular study, each subject was given a post-test survey with four questions. The questions indicated subject enjoyment, perception regarding the usefulness of each agent in a physical therapeutic session, difficulty of action recognition, and difficulty of repeating the recognized action. Table 10 lists the specific questions from this survey.

A total of 11 adults participated in the study which was directed in the same

Table 11: Results of the Post-test Survey

Rated from 1 to 5		
	Mean Response	Stdv
Q1		
simulated	2.9	1.0
embodied	4.2	1.1
Rated from 1 to 10		
Q2		
simulated	3.4	2.3
embodied	2.7	2.3
Q3		
simulated	3.3	2.2
embodied	3.1	2.7
Yes or No Response		
Q4	Total Yes	Total No
simulated	6	5
embodied	11	0

manner as the previously described study in Section 5.3. A T-Test between-subjects indicated that the participants enjoyed their interaction with the embodied agent more than with the simulated agent ($p < 0.01$). Table 11 shows the results of each question from the survey. Also, all 11 participants indicated that the embodied agent would be beneficial in physical therapy, whereas only six participants indicated that the simulated agent would be beneficial in the same capacity. There was no significant difference in regards to the difficulty of recognizing and repeating the actions of each agent. The results of this study indicated that subjects preferred to interact with an embodied agent rather than a simulated agent. However, we postulated as to whether or not the overall desire to interact with the embodied agent as opposed to the simulated agent had an effect on a subject's ability to comply with the instructions given via a particular agent. Chronologically, the following experiment was conducted.

5.3 Experimental Protocol

Using the same approach outline in Chapters 2 and 3 for calculating the kinematic metrics, testing took place in the Motor Development Lab in the Division of Physical Therapy at Georgia State University under the direction of Dr. Yu-Ping Chen. Each subject was given an IRB approved consent form informing him or her of the testing procedure. After consent was given, the testing procedure began. The procedure was conducted as follows:

Task 1:

Subjects were shown an on-screen simulated robot. Subjects were told to watch the movement of the simulated robot and to repeat its movement once it has stopped.

Task 2:

Subjects were shown an embodied robot. Subjects were told to watch the movement of the embodied robot and to repeat its movement once it has stopped.

Each task consisted of three different actions: shoulder abduction, elbow flexion, and reaching. The starting and ending position of each movement are shown in Figures 55, 56, and 57 respectively. The testing procedure was a between-subjects study with each subject interacting with both agents in a random order (i.e. the simulated agent performed the action first or the embodied agent performed the action first). Each participant was given a post-test survey in order to determine his or her perception of the overall experience.

5.3.1 Apparatus

All participants stood directly in front of the simulated agent, depicted on a projector screen, and the embodied agent. A 100Hz six-camera Vicon (370) Motion Capture System (VMCS) recorded position data of the subjects' joints. Three small reflective

(a) (b)

Figure 55: Illustrations of the a)starting and b)ending positions of the shoulder abduction exercise.

(a) (b)

Figure 56: Illustrations of the a)starting and b)ending positions of the elbow flexion exercise.

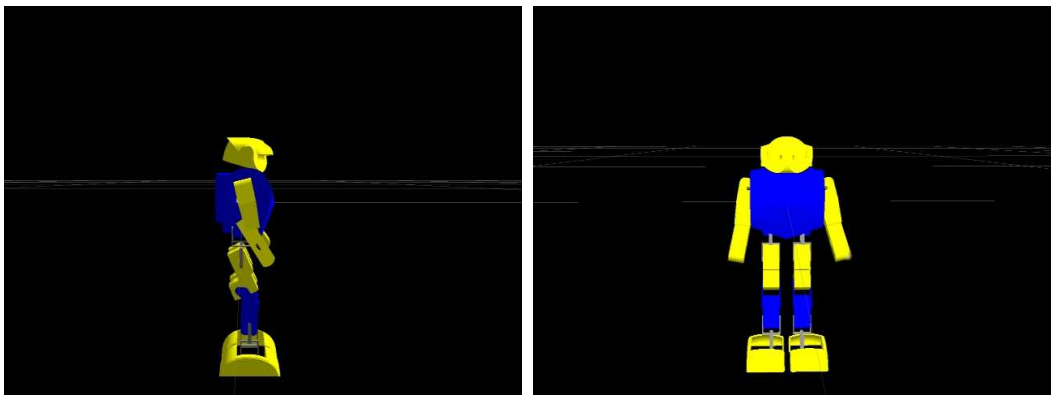
(a) (b)

Figure 57: Illustrations of the a)starting and b)ending positions of the reaching exercise.

markers (9 mm in diameter) were attached to each subject's scapula (shoulder), lateral epicondyle of the humerus (elbow), and ulnar styloid process (wrist). VMCS cameras were placed around the participant at a distance of 5m in order to track the reflective markers simultaneously. The VMCS cameras in relation to the testing were initially calibrated by using a calibration rod. In addition to the kinematic data, movements were recorded throughout the collection session with a Microsoft Kinect®, which, again, allows collection of depth information for the purpose of comparing it with the kinematic data. Thus, motion was captured via two methods.

5.3.2 Simulated Agent

The simulated agent depicted a model version of the Manoi humanoid introduced by Kyosho and was projected onto a projector screen. The simulated agent was constructed in-house by the Human Automation Systems (HumAnS) Laboratory at the Georgia Institute of Technology using the Robot Operating System (ROS) in conjunction with AutoCAD (a software application for computer-aided design and drafting in both 2D and 3D), Blender (a 3D content creation suite), and Rviz (a



(a) (b)

Figure 58: a)Side-view and b)front-view of the simulated agent.

Figure 59: Illustration of the embodied agent (Darwin-OP).

graphics simulator used to depict model robots). When given the command via a laptop, the agent performed a prescripted upper-limb motion for one-half a repetition. Figure 58 is an illustration of the simulated agent that was used for testing.

5.3.3 Embodied Agent

The embodied agent that was utilized was the Darwin-OP (Dynamic Anthropomorphic Robot with Intelligence - Open Platform) [93]. It is 455mm (17.9in) in height and weighs 2.8 kg (6.2 lbs). It is equipped with Robotis Dynamixel motors, a FitPC for computing, and LiPo batteries for power. The Darwin-OP was also prescripted to

Figure 60: Illustrations of the test setup.

Table 12: Survey Questionaire from Pilot Study

Q1	I could see all the movements from Darwin very well.
Q2	The movements that Darwin requested were too hard.
Q3	Darwin was so attractive that I lost track of time.
Q4	I would like to play with Darwin more often.

perform an upper-limb motion for one-half a repetition using a ROS package designed in-house by the Human Automation Systems (HumAnS) Laboratory at the Georgia Institute of Technology. Figure 59 is an illustration of the embodied agent that was used for testing. The entire testing setup, which includes each apparatus and both agents, can be seen in Figure 60.

5.3.4 Post-Test Survey

In order to determine whether or not human perception has any significance on a subject's ability to adhere to proper exercises in this specific scenario, a post-test survey was given regarding the overall session. The questionnaire consisted of four questions regarding the ease of recognizing movements, difficulty of mimicking actions, physical appeal, and overall enjoyment. Table 12 lists the specific questions that pertain to the embodied agent. The evaluation used a 5-point Likert Scale: Completely Disagree = 1; Slightly Disagree = 2; Neutral = 3; Slightly Agree = 4; Completely Agree = 5.

Table 13: Average Percentage Error for Shoulder Abduction Kinematic Data for VMCS versus Vision Algorithm

Parameter	Error %	Stdv
AROM	4.17	6.97
PAV	4.00	0.57
NJS	6.30	0.24

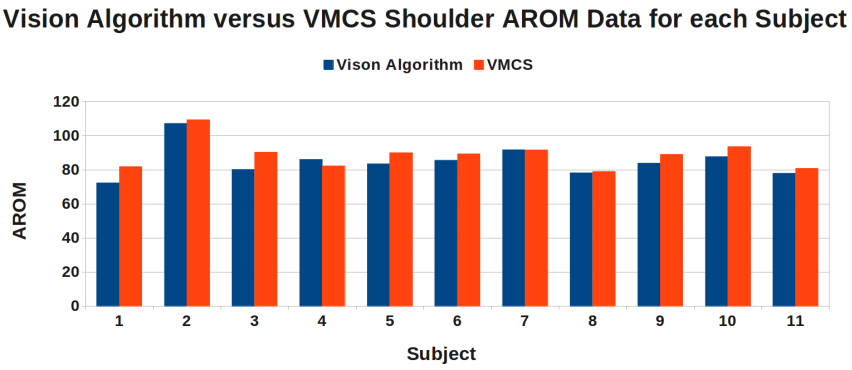


Figure 61: Comparison between the Vision Algorithm and VMCS of each subject's AROM for shoulder abduction.

5.4 Results of the Study

Eleven healthy young adults between the ages of 19 and 22 performed each task – once per agent mimicking. The first step was to show the legitimacy of the image processing approach for calculating the kinematic metrics. This was achieved by calculating the error between the VMCS and the vision algorithm data for each subject's individual movements. In other words, although the capturing systems are diferent, the movements are exactly the same. The average errors of all subjects for each kinematic metric per exercise are shown in Tables 13, 14, and 15. Figures 61 - 70 show each subject's respective data for each kinematic metric per exercise.

We were also able to obtain data from one child subject who performed a shoulder abduction and elbow flexion exercise. The results from the child subject, shown in Table 16, are still within the error performance measures as those for the adult subjects. This provides an initial confidence that our approach is indeed a viable one

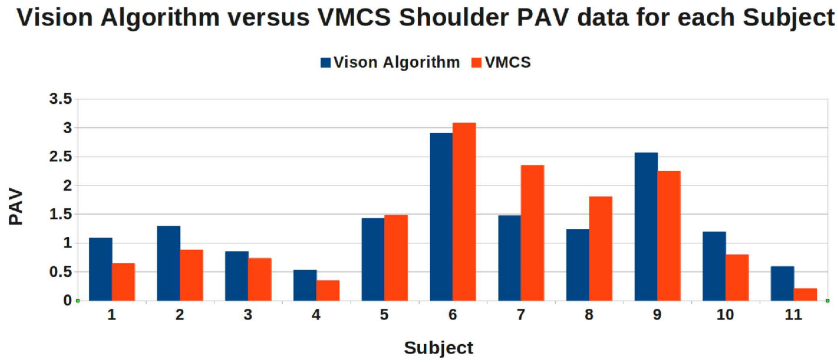


Figure 62: Comparison between the Vision Algorithm and VMCS of each subject's PAV for shoulder abduction.

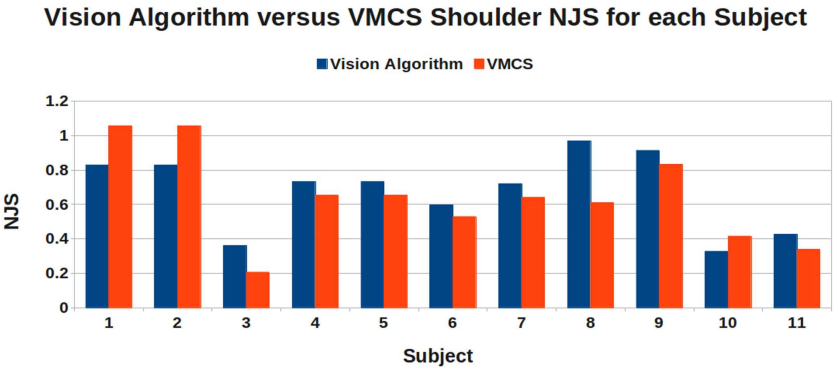


Figure 63: Comparison between the Vision Algorithm and VMCS of each subject's NJS for shoulder abduction.

Table 14: Average Percentage Error for Elbow Flexion Kinematic Data for VMCS versus Vision Algorithm

Parameter	Error %	Stdv
AROM	4.03	7.22
PAV	4.97	0.41
NJS	6.93	0.22

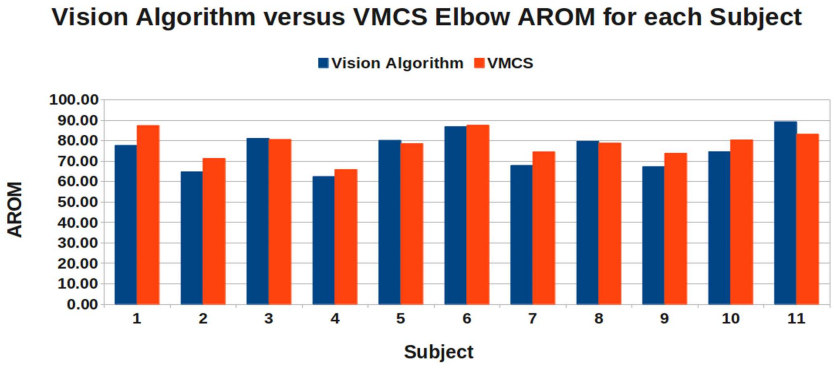


Figure 64: Comparison between the Vision Algorithm and VMCS of each subject's AROM for elbow flexion.

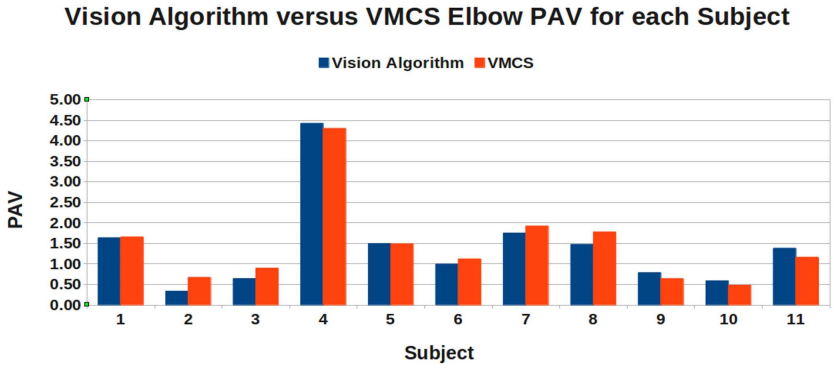


Figure 65: Comparison between the Vision Algorithm and VMCS of each subject's PAV for elbow flexion.

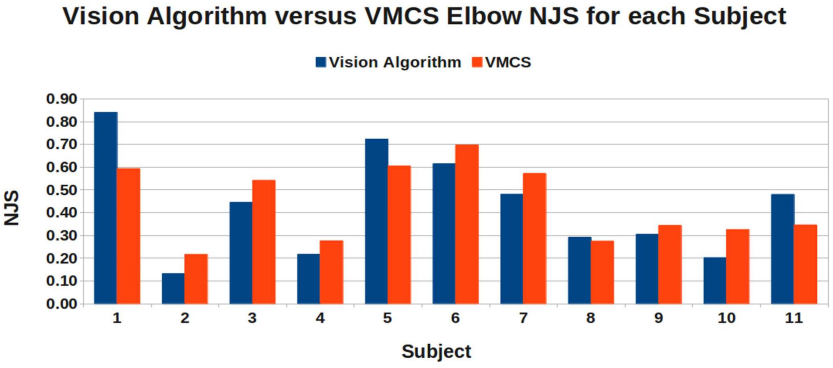


Figure 66: Comparison between the Vision Algorithm and VMCS of each subject's NJS for elbow flexion.

Table 15: Average Percentage Error Reaching Kinematic Data for VMCS versus Vision Algorithm

Parameter	Error	Stdv
MT	2.54	1.21
TD	19.69	6.46
PAV	12.06	0.29
MU	0.00	0.00
NJS	9.22	0.12

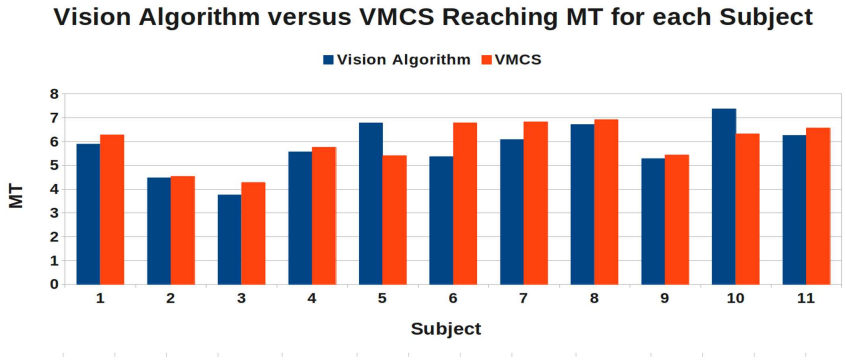


Figure 67: Comparison between the Vision Algorithm and VMCS of each subject's MT for reaching.

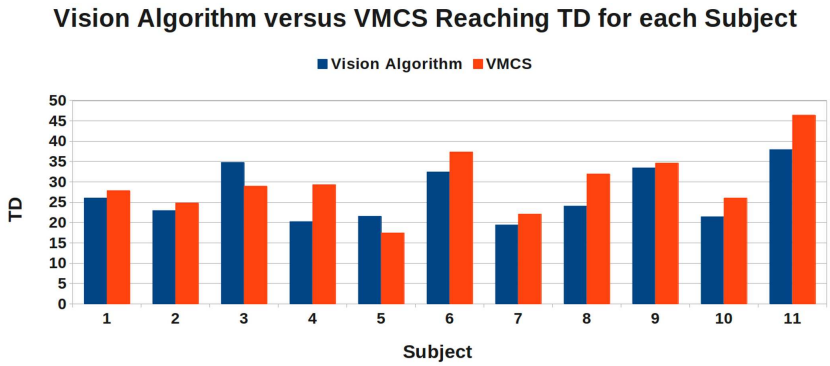


Figure 68: Comparison between the Vision Algorithm and VMCS of each subject's TD for reaching.

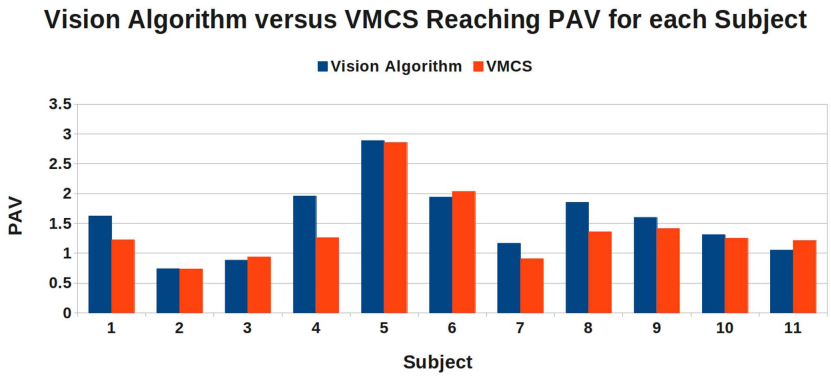


Figure 69: Comparison between the Vision Algorithm and VMCS of each subject's PAV for reaching.

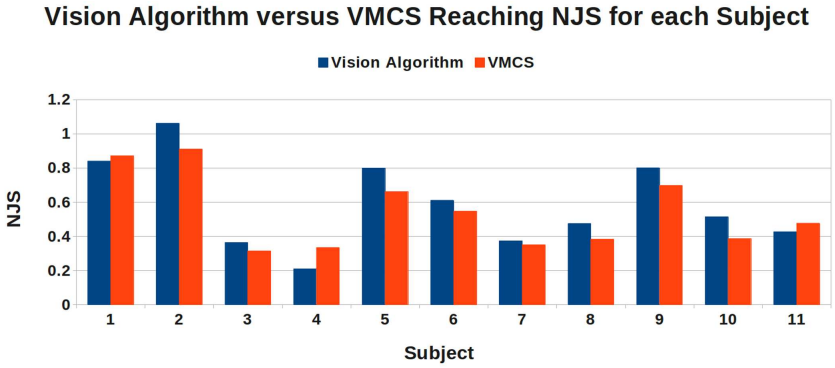


Figure 70: Comparison between the Vision Algorithm and VMCS of each subject's NJS for reaching.

Table 16: Child Percentage Error for Shoulder Abduction and Elbow Flexion Kinematic Data for VMCS versus Vision Algorithm

Parameter	Shoulder %	Elbow %
AROM	7.13	9.07
PAV	11.25	13.53
NJS	16.83	18.06

Table 17: Average Percentage Error for Shoulder Abduction Kinematic Data for Simulation versus Embodied Interaction

Parameter	Simulation %	Stdv	Embodied %	Stdv
AROM	16.59	24.33	9.08	10.84
PAV	57.82	2.11	57.21	2.15
NJS	29.08	0.29	22.81	0.41

towards quantifying upper-arm rehabilitation metrics for children through interaction with a humanoid robot.

In order to determine proper correlation with regard to interaction with the simulated agent versus the embodied agent, the average error of all subjects for the kinematic metric data for each exercise was calculated for each interaction only using the image processing method. The ground truth for the metrics was obtained from the prescripted motion that each agent mimicked during testing (as stated earlier), which stemmed from the pre-recording of an expert's kinematic metric data obtained via the same method. Tables 17, 18, and 19 are presentations of the resulting data. Figures 71 - 80 show each subject's respective data for each kinematic metric per exercise in comparison to the human expert's kinematic metrics.

5.4.1 Discussion

The magnitudes of error between the VMCS and image processing data versus that of the simulated versus embodied agents are drastically different. It is logical that when comparing the VMCS exercise data of Subject X with the image processing data of

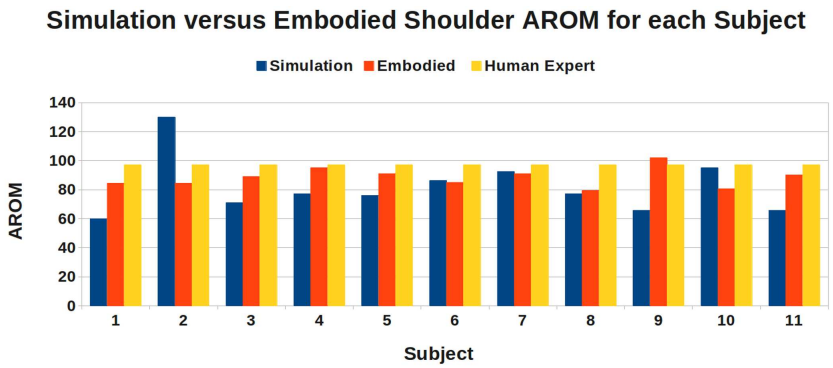


Figure 71: Comparison between the simulation and embodied agent interaction of each subject's AROM for shoulder abduction.

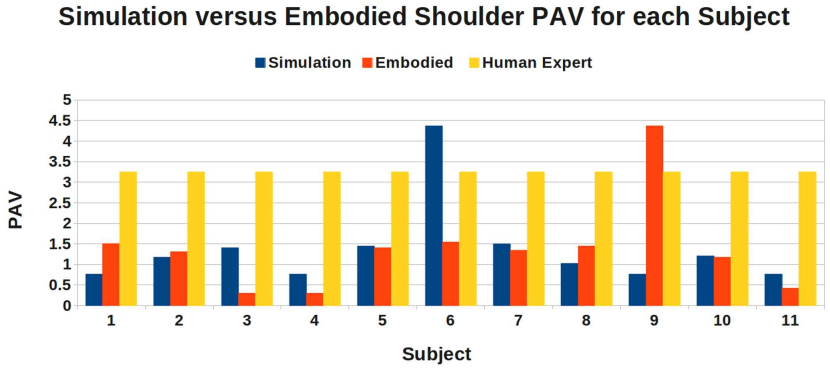


Figure 72: Comparison between the simulation and embodied agent interaction of each subject's PAV for shoulder abduction.

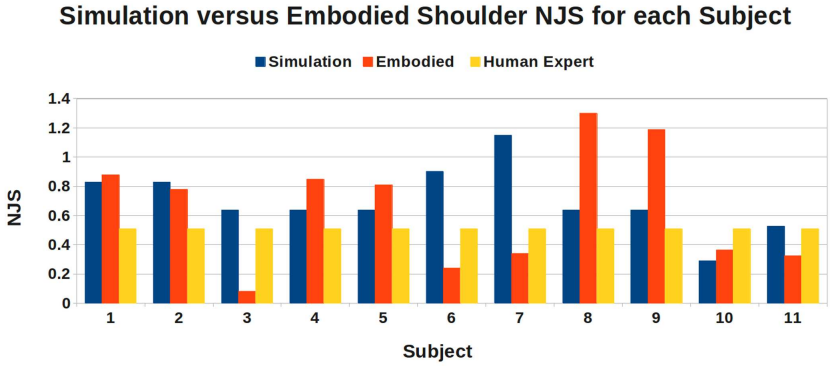


Figure 73: Comparison between the simulation and embodied agent interaction of each subject's NJS for shoulder abduction.

Table 18: Average Percentage Error for Elbow Flexion Kinematic Data for Simulation versus Embodied Interaction

Parameter	Simulation %	Stdv	Embodied %	Stdv
AROM	30.34	33.81	16.13	23.62
PAV	49.58	1.75	44.95	1.74
NJS	83.14	0.37	62.63	0.25

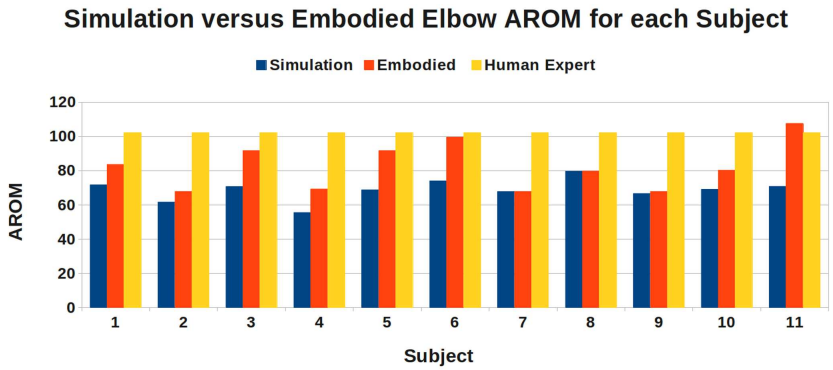


Figure 74: Comparison between the simulation and embodied agent interaction of each subject's AROM for elbow flexion.

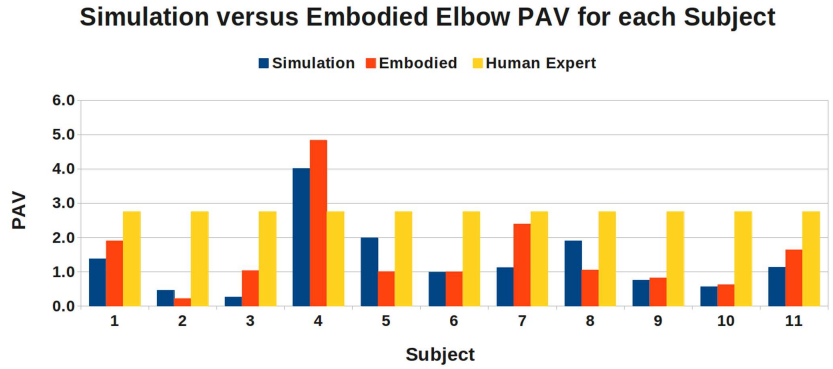


Figure 75: Comparison between the simulation and embodied agent interaction of each subject's PAV for elbow flexion.

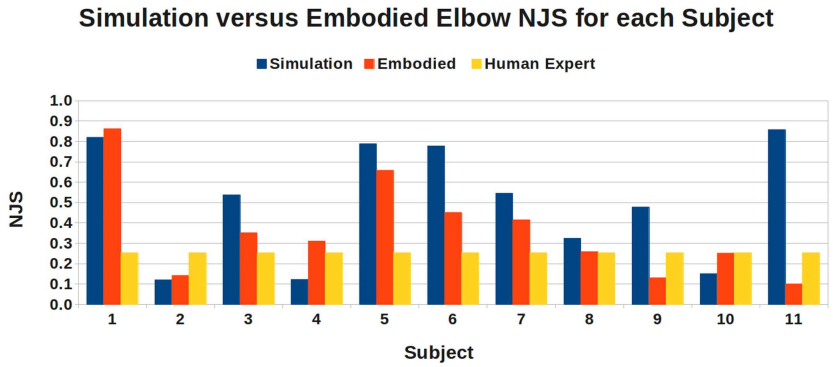


Figure 76: Comparison between the simulation and embodied agent interaction of each subject's NJS for elbow flexion.

Table 19: Average Percentage Error for Reaching Kinematic Data for Simulation versus Embodied Interaction

Parameter	Simulation %	Stdv	Embodied %	Stdv
MT	14.70	1.87	31.37	1.14
TD	9.26	7.80	6.72	6.11
PAV	7.77	0.63	18.38	1.07
MU	0.00	0.00	0.00	0.00
NJS	17.27	0.26	5.29	0.33

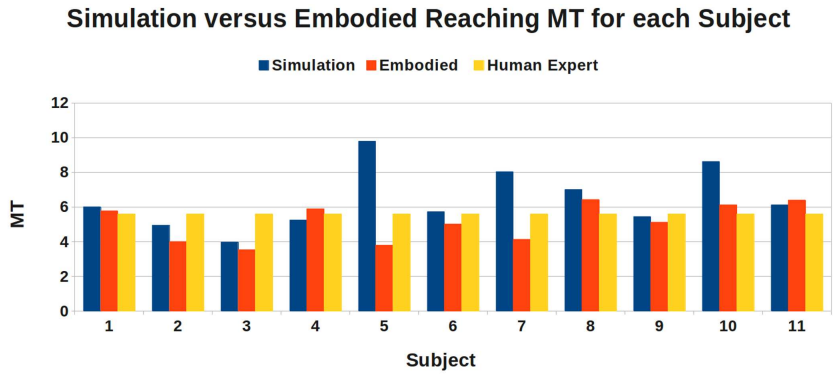


Figure 77: Comparison between the simulated and embodied agent interaction of each subject's MT for reaching.

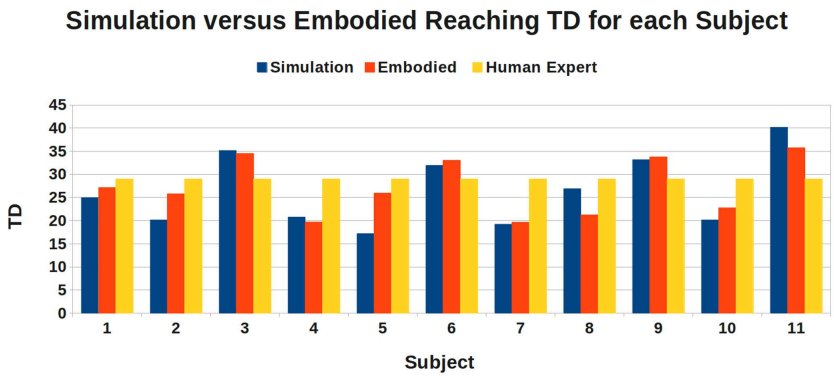


Figure 78: Comparison between the simulated and embodied agent interaction of each subject's TD for reaching.

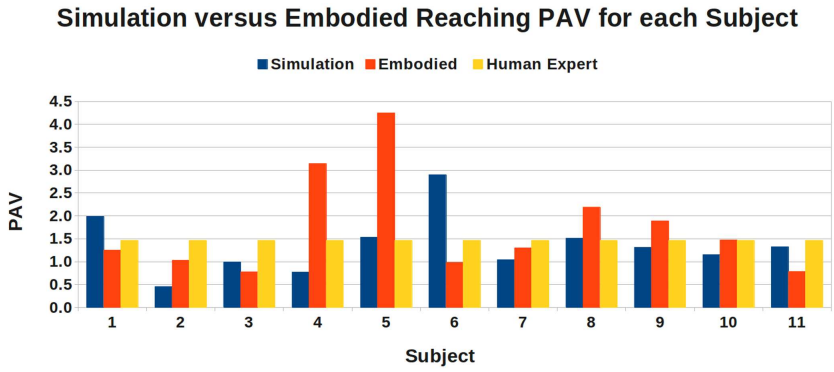


Figure 79: Comparison between the simulated and embodied agent interaction of each subject's PAV for reaching.

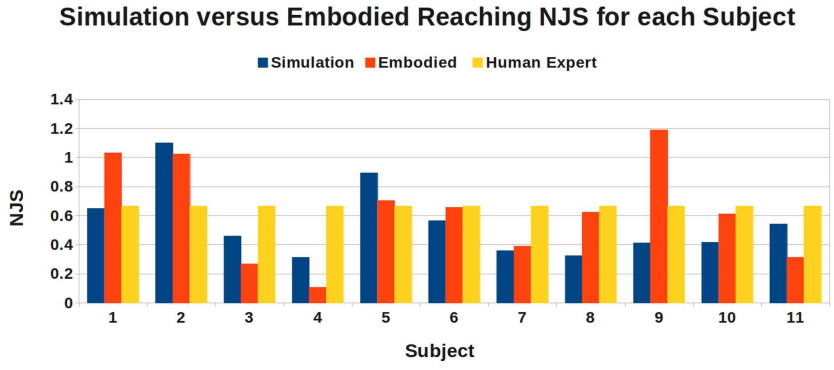


Figure 80: Comparison between the simulated and embodied agent interaction of each subject's NJS for reaching.

the same subject (Subject X), the magnitude of the error should not be terrible high. In other words, the capturing systems are different, but the movements are exactly the same. This is considered the ground truth data set. On the other hand, it is also logical that when comparing the image processing data of Subject X with the image processing data of Subject Y (a human expert's kinematic data set), the magnitude of the error will most likely be higher. Stated differently, one human can not identically mimic the actions of another human without some degree of variance. The capturing systems are the same, but the movements, while very similar, are not identical.

Furthermore, each subject was only asked to mimic the overall motion of each agent. Subjects were not told that they should also try to mimic the movement speed and displacement of each agent. While, each subject did indeed perform the basic actions that were demonstrated, most subjects did not focus on certain intricacies. More specifically, the subjects were not attentive of the exact height and speed of the demonstrated movements that were created from a human expert's kinematic data set, which, based upon observation, effected the successive kinematic measures. Perhaps more important in this scenario are simply the magnitudes of difference between the interactions with the simulated and embodied agents.

With the exceptions of certain metrics from the reaching kinematics, the subjects adhered better to the movements that were demonstrated by the embodied agent. The results of the reaching data seem to indicate that subjects relate better to the speed of movement with the simulated agent than with the embodied agent. Perhaps the most interesting result is that of the NJS. It appears that the amount of jerk was more closely related to the embodied interaction than with the simulated counterpart.

5.4.2 Survey Analysis

The post-test survey indicated that it was relatively simple for the subjects to recognize the movements of the embodied agent. Based on the responses from question

Table 20: Statistical Results of the Post-test Survey

Options: Completely Disagree = 1; Slightly Disagree = 2; Neutral = 3; Slightly Agree = 4; Completely Agree = 5		
Question	Mean Response	Stdv
1	5.0	0.0
2	1.2	0.6
3	3.1	1.2
4	3.9	1.0

1, all participants agreed that they could see the movements of the embodied agent well. The responses from question 2 to indicated that each participant could easily mimic movements from the embodied agent. The responses from questions 3 and 4 indicate that the majority of the participants enjoyed interacting with the embodied again and would be willing to continue interacting with it.

In the free response section, many subjects explicitly stated that it was more difficult to visualize movements that were being shown by the simulated agent on the projection screen. Those statements gave insight to subject's responses regarding the ease of following the embodied agent, noting that the requested movements were not too difficult. Also, the physical appeal of the embodied agent seemed to captivate the subjects. One subject stated, "Darwin was awesome! It's like a real-life 3-D version of the Kinect." This also gave insight to the overall enjoyment of the experience, which was a positive outcome. The resulting averages and standard deviations are shown in Table 20.

5.5 Conclusions

It has been shown that kinematic metrics for different upper-limb exercises can be extracted with a simple depth camera. Furturemore, it has also been shown that proper compliance to exercises does have dependence upon the agent with which a patient may be interacting. It is suspected that the ability of humans to decipher

movements in general may lead to some degree of proper mimicking regardless of the agent. It is possible, however, that more difficult exercises would require the instruction of an agent capable of providing a real-world perspective rather than a virtual view, such as the perspective provided by the embodied agent. Perhaps just as important are the desires of the patient to interact with a specific agent. Increasing the desire to perform exercises that are requested by a physical therapist by incorporating interactive agents could indeed result in higher levels of engagement and decreases in recovery time. The potential for increases in attentiveness and engagement could aid both adult and child subjects by making the therapeutic process seem less tedious and more exciting.

CHAPTER VI

CONCLUSIONS AND DISCUSSION

Historically, assistive therapy has been centered around people aiding people for the purpose of providing a better way of life. Whether a person physically aids someone by assisting him or her with walking or gives cognitive assistance by teaching him or her methods of memorization, assistive therapy has been and will continue to be implemented in various ways. The motivation is to allow those with impairments to have the same quality of life as those without impairments.

The progression of this field has been to move away from human assistance to machine assistance, albeit improving current methods and decreasing the amount of manual labor. Utilizing these devices has had a great impact on people with impairments. The consistent decrease in the cost of obtaining imaging devices and the increase in quality has enabled researchers to develop new and efficient modes of exploring and analyzing the world. The work presented in this research has added to this progression. By coupling several computer vision techniques with robotics, it has been shown that a novel method of tracking a human subject and analyzing his or her movements with regards to specific exercises and metrics can be achieved. It has also been shown that many humans believe that robotics can have a positive influence in the field of rehabilitation therapy. As such, it is necessary that therapists and researchers in both medical and engineering fields continue to collaborate in order to provide novel methods of rehabilitation therapy via robotic assistance.

6.1 Limitations

When considering the approach presented here for gathering physical therapeutic data, one has to be knowledgeable of the limitations or potential concerns. Some of

those limitations were outlined in Chapter 4 Section 4.3. In addition, many physical therapists also use sensation or force-feedback as a method of assessing the patient's current state and determining the effects of the physical therapy sessions. Since this algorithm focuses on non-contact assessments, it will not be possible to gather important information obtained from sensation measurements. This limitation in-turn creates a concern as to whether or not this method will produce comparable effects to robotic rehabilitation devices that incorporate force-feedback with subtle, unrecognizable cues to the patient, changes that induce motor movements that the patient believes he or she is incapable of achieving due to injury or disability. It is quite possible, however, that this approach will have positive results in that it will provide a motivating factor for individuals during the therapeutic sessions. Thus, with an increase in motivation and attention, it is hypothesized that this will indeed produce comparable results to robotic rehabilitation devices that incorporate force-feedback. The aim, though, is to provide multiple methods of assessing a patient's progress. Methods that utilize force-feedback are necessary for assessing a patient's strength or lack thereof (e.g. hemiparesis), while non-contact assessments can be used in less severe cases (e.g. a previously dislocated joint).

While it is possible to achieve rotation via the Principle Component Analysis calculation, described in Chapter 2 Section 2.6, even for motions where the subject is directly facing the camera, this point of view is not analogous to a typical therapeutic assessment. Meaning, physical therapists assess a patient's Active Range of Motion using a goniometer from beside the patient rather than directly in front of him or her. When monitoring the patient from a frontal view, it becomes difficult to determine the angular variation between the upper and lower line segments of the arm during the exercise. The Principle Component Analysis will indeed rotate the image to its proper form; however, the segmented movement is no longer seen as an angular motion. Rather it appears as a vertical or horizontal shift. Figures 81 and 82 in

Appendix A illustrate this scenario.

The typical reaching kinematic metrics (Movement Time and Total Displacement) can still be calculated, however, because they are not solely dependent upon the Active Range of Motion or angular positioning. As stated in Chapter 3 Section 3.3.3, the Peak Velocity can be calculated by dividing the Total Displacement by the time at each frame and taking the maximum value. The Movement Units and Normalized Jerk Score would still be functions of the previously calculated velocity values. This is all stated to indicate that certain parameters (MT, TD, PV, MU, and NJS) are feasible from a frontal view, while other parameters (AROM and PAV) are not.

The last limitation addresses the issue of processing speed. Currently, due to the low processing capability of the Darwin-OP and the high processing demands of the Microsoft Kinect®, it is impractical to perform real-time data collection and analysis on-board the Darwin-OP. The Microsoft Kinect® operates at 2.4GHz, whereas the Darwin-OP operates at 1.2GHz. Thus the data has to be retrieved and processed on a separate computer with sufficient operating speed.

6.2 Future Direction

There are several future directions that this research should take. In the immediate future, more child subjects should be tested across a long-term study. Ideally, a long-term study that utilized both neurotypical and non-neurotypical children would be beneficial. This study could give insight regarding the effectiveness of using robotics as physical aids for children with and without neurotypical disorders.

In addition, it would also be beneficial to continue the study of human compliance. This could be achieved in one of two ways. First, the study could be continued in an outpatient rehabilitation facility where the robotic platform could continue to be utilized. Second, due to cost constraints, specifically referring to the current cost of humanoid robots, patients could be allowed to take the Microsoft Kinect® to their

home in exchange for agreeing to record their motions while interacting with a video instructor (e.g. human recording or simulated model). As stated previously, proper compliance is an important factor in the recovery process. Thus, continuing this study could result in more effective measures, longer engagement during interaction, and decreased recovery time.

Finally, a real-time integrated system should be the final goal. Ideally, the system would be integrated onto a robotic platform capable performing its movements, video capture, and image processing completely on-board while interacting with human subjects autonomously. This would require the design and implementation of a robot with much higher processing power than most of the current systems that are available. While it is feasible to perform the data collection and processing on a separate computer, the next step should adhere to the logical future of robotics, which is complete autonomy and real-time interaction.

6.3 Final Thoughts

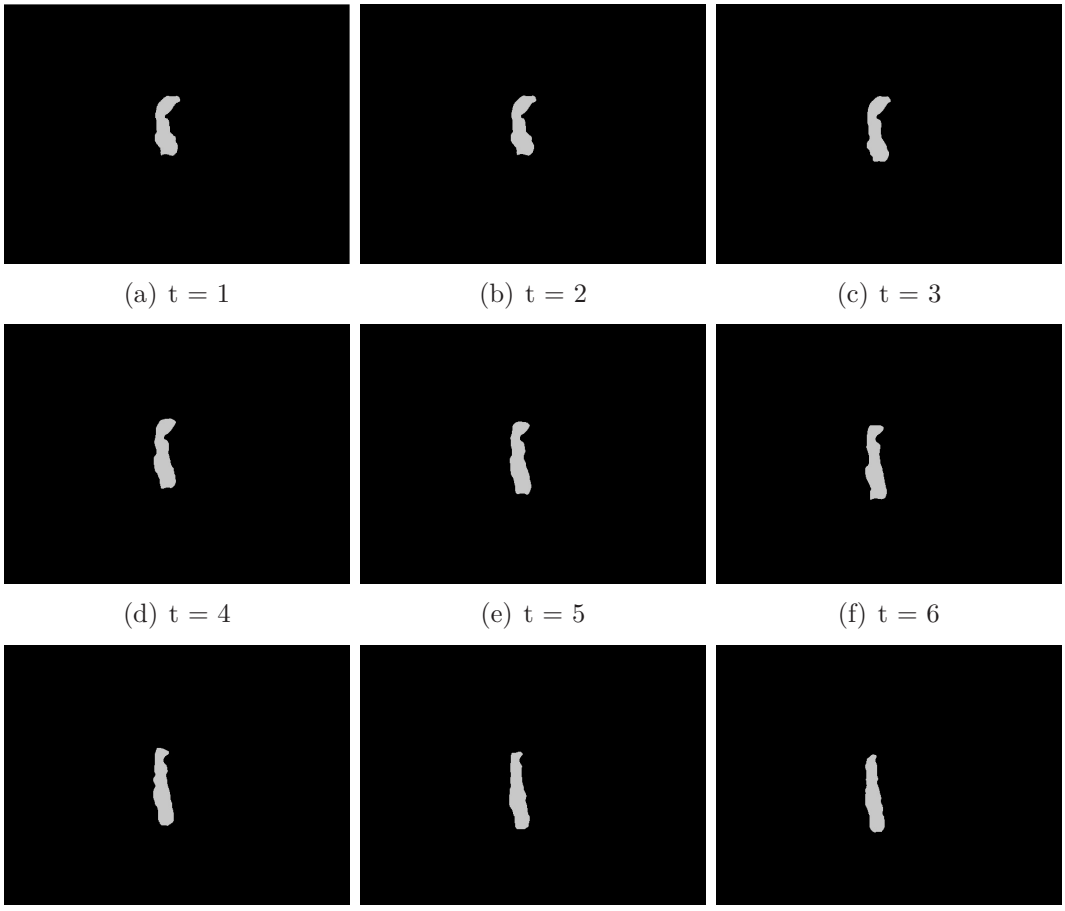
The future of robotics in health-care has great promise. As their use in the field continues to progress, it becomes more apparent for the need to integrate multiple aspects of interaction and reasoning. The research presented here has shown a few of the different facets with which robotics have been implemented in the health-care field. Future health-care robotics will be equipped to provide even more analyses, such as interpreting discomfort levels as provided by vital signs or facial expressions.

With regards to upper-limb rehabilitation, various platforms such as those mentioned in the literature review have already shown positive results both in laboratory settings and in outpatient therapeutic sessions. It is believed that the future direction of robots geared towards aiding individuals with upper-limb impairments lies within cross-discipline studies and collaborations, those that include Biology, Computer Science, Electrical Engineering, and Psychology to name a few. By incorporating the

vast information from multiple disciplines, robotic designs will be able to provide a robust and thorough analysis of a patient's current impairment and progress throughout his or her therapeutic sessions, both during outpatient therapy and at home. The research presented here is a stepping stone for the continued progression of robotics in the health-care sector and the general acceptance of robotics in society as a whole.

APPENDIX A

FIGURES ILLUSTRATING FRONTAL VIEW SEGMENTATION

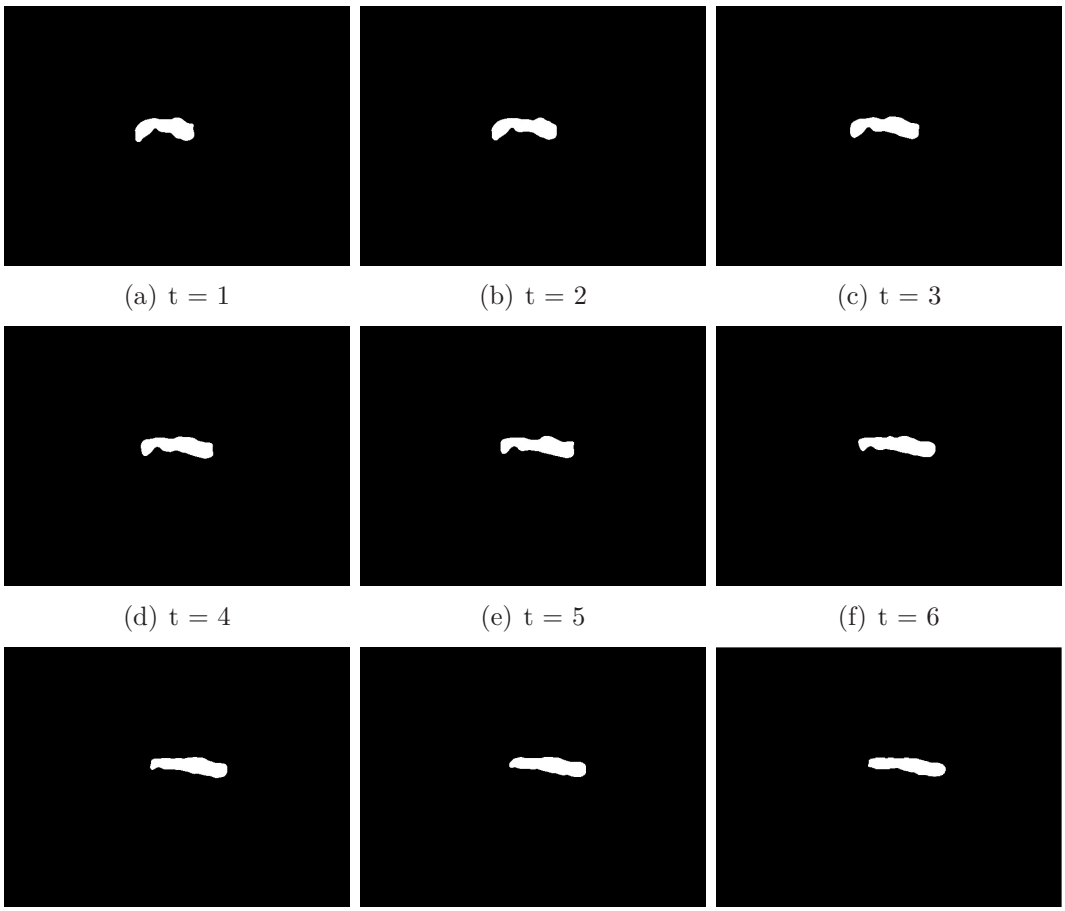


(a) t = 1 (b) t = 2 (c) t = 3

(d) t = 4 (e) t = 5 (f) t = 6

(g) t = 7 (h) t = 8 (i) t = 9

Figure 81: Illustration of segmenting the sequential movements of a subject from a frontal view. The variable 't' represents time in seconds.



(a) t = 1 (b) t = 2 (c) t = 3

(d) t = 4 (e) t = 5 (f) t = 6

(g) t = 7 (h) t = 8 (i) t = 9

Figure 82: Illustration of the resulting PCA projected images after segmenting the sequential movements of a subject from a frontal view. The variable 't' represents time in seconds.

REFERENCES

[1] Alberts, J., Saling, M., Adler, C., and Stelmach, G., "Disruptions in the reach-to-grasp actions of parkinson's patients," *Experimental brain research*, vol. 134, no. 3, pp. 353–362, 2000.

[2] Apewokin, S., Valentine, B., Forsthoefel, D., Wills, L., Wills, S., and Gentile, A., "Embedded real-time surveillance using multimodal mean background modeling," *Embedded Computer Vision*, pp. 163–175, 2009.

[3] Asghari Oskoei, M. and Hu, H., "Myoelectric control systems–A survey," *Biomedical Signal Processing and Control*, vol. 2, no. 4, pp. 275–294, 2007.

[4] Au, A. and Kirsch, R., "EMG-based prediction of shoulder and elbow kinematics in able-bodied and spinal cord injured individuals," *Rehabilitation Engineering, IEEE Transactions on*, vol. 8, no. 4, pp. 471–480, 2000.

[5] Beer, R., Chiel, H., and Drushel, R., "Using autonomous robotics to teach science and engineering," *Communications of the ACM*, vol. 42, no. 6, pp. 85–92, 1999.

[6] Blow, M., Dautenhahn, K., Appleby, A., Nehaniv, C., and Lee, D., "The art of designing robot faces: Dimensions for human-robot interaction," in *Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction*, p. 332, ACM, 2006.

[7] Bobick, A. F. and Davis, J. W., "The recognition of human movement using temporal templates," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, pp. 257–267, March 2001.

[8] Box, G., Jenkins, G., Reinsel, G., and others, *Time series analysis: forecasting and control*, vol. 16. Holden-day, San Francisco, 1976.

[9] Brenton, H., Gillies, M., Ballin, D., and Chatting, D., "The uncanny valley: does it exist," in *proc HCI Annu Conf: workshop on human-animated character interaction, Edinburgh*, 2005.

[10] Brewer, B., McDowell, S., and Worthen-Chaudhari, L., "Poststroke upper extremity rehabilitation: a review of robotic systems and clinical results," *Topics in Stroke Rehabilitation*, vol. 14, no. 6, pp. 22–44, 2007.

[11] Brisben, A., Safos, C., Lockerd, A., Vice, J., and Lathan, C., "The CosmoBot^{TM} System: Evaluating its Usability in Therapy Sessions with Children Diagnosed with Cerebral Palsy," 2005.

[12] Brokaw, E., Murray, T., Nef, T., and Lum, P., "Retraining of interjoint arm coordination after stroke using robot-assisted time-independent functional training," *Journal of Rehabilitation Research and Development*, vol. 48, 2010.

[13] Brooks, D. and Howard, A., "Quantifying Upper-Arm Rehabilitation Metrics for Children through Interaction with a Humanoid Robot," *Applied Bionics and Biomechanics*, 2011.

[14] Brooks, D. and Howard, A. M., "Upper-limb Rehabilitation and Evaluation of Children using a Humanoid Robot," in *ICMI-MLMIâĂŹ09 Workshop on Child, Computer and Interaction*, November 2009.

[15] Burgar, C., Lum, P., Shor, P., and der Loos, M. V., "Development of robots for rehabilitation therapy: the palo alto va/stanford experience," *J Rehabil Res Dev.*, vol. 37, pp. 663–673, Nov-Dec 2000.

[16] Butler, E., Ladd, A., Louie, S., LaMont, L., Wong, W., and Rose, J., "Three-dimensional kinematics of the upper limb during a reach and grasp cycle for children," *Gait & posture*, vol. 32, no. 1, pp. 72–77, 2010.

[17] Cameirão, M., Badia, S., Oller, E., Verschure, P., and others, "Neurorehabilitation using the virtual reality based rehabilitation gaming system: methodology, design, psychometrics, usability and validation," *Journal of neuroengineering and rehabilitation*, vol. 7, no. 1, p. 48, 2010.

[18] Campbell, L. and Bobick, A. F., "Recognition of human body motion using phase space constraints," in *Proc. Int'l Conference Computer Vision*, pp. 624–630, 1995.

[19] Canny, J., "A computational approach to edge detection," *Readings in Computer Vision: Issues, Problems, Principles and Paradigms*, vol. 8, no. 6, pp. 184–203, 1986.

[20] Carroll, D., "A quantitative test of upper extremity function," *Journal of Chronic Diseases*, vol. 18, no. 5, pp. 479–491, 1965.

[21] Casadio, M., Giannoni, P., Morasso, P., and Sanguineti, V., "A proof of concept study for the integration of robot therapy with physiotherapy in the treatment of stroke patients," *Clinical Rehabilitation*, vol. 23, no. 3, p. 217, 2009.

[22] CDC, "Cdc–arthritis–data and statistics." http://www.cdc.gov/arthritis/data_statistics.htm, accessed Feb 2012.

[23] CDC, "Cdc–dhdsp–stroke facts." http://www.cdc.gov/stroke/facts.htm, accessed Feb 2012.

[24] Chan, A. and Englehart, K., "Continuous myoelectric control for powered prostheses using hidden Markov models," *Biomedical Engineering, IEEE Transactions on*, vol. 52, no. 1, pp. 121–124, 2005.

[25] Chan, J., Leung, H., and Poizner, H., "Rehabilitation engineering-correlation among joint motions allows classification of parkinsonian versus normal 3-d reaching," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 18, no. 2, p. 142, 2010.

[26] Chang, J., Wu, T., Wu, W., and Su, F., "Kinematical measure for spastic reaching in children with cerebral palsy," *Clinical Biomechanics*, vol. 20, no. 4, pp. 381–388, 2005.

[27] Chiang, J., Wang, Z., and McKeown, M., "Hidden Markov Multivariate Autoregressive (HMM-mAR) Modeling Framework for Surface Electromyography (sEMG) Data," in *Engineering in Medicine and Biology Society, 2007. EMBS 2007. 29th Annual International Conference of the IEEE*, pp. 4826–4829, IEEE, 2007.

[28] Chum, O., *Two-view geometry estimation by random sample and consensus*. PhD thesis, Czech Technical University, 2005.

[29] Cleveland Medical Devices, I., "BioCapture, BioRadio 150." http://www.clevemed.com/BioCapture/hardware.shtml, 2008.

[30] Colombo, R., Pisano, F., Micera, S., Mazzone, A., Delconte, C., Carrozza, M., Dario, P., and Minuco, G., "Upper limb rehabilitation and evaluation of stroke patients using robot-aided techniques," in *ICORR 2005 Conference Proceedings*, pp. 515–518, June 28 - July 1 2005.

[31] Cooley, E. and Morris, R., "Attention in children: A neuropsychologically based model for assessment," *Developmental Neuropsychology*, vol. 6, no. 3, pp. 239–274, 1990.

[32] Cormen, T., Leiserson, C., and Rivest, R., "0-262-03141-8," 1998.

[33] Cromwell, F., "Occupational therapists manual for basic skills assessment: primary prevocational evaluation," *Pasadena, CA: Fair Oaks Printing*, pp. 29–31, 1965.

[34] Dautenhahn, K., "Robots as social actors: Aurora and the case of autism," in *Proc. CT99, The Third International Cognitive Technology Conference, August, San Francisco*, pp. 359–374, 1999.

[35] Dautenhahn, K. and Billard, A., "Games children with autism can play with robota, a humanoid robotic doll," *Universal Access and Assistive Technology*, pp. 179–190, 2002.

[36] Dong, P., Brankov, J., Galatsanos, N., Yang, Y., and Davoine, F., "Digital watermarking robust to geometric distortions," *Image Processing, IEEE Transactions on*, vol. 14, no. 12, pp. 2140–2150, 2005.

[37] Duda, R. and Hart, P., "Use of the Hough transformation to detect lines and curves in pictures," *Communications of the ACM*, vol. 15, no. 1, pp. 11–15, 1972.

[38] Duncan, P., Propst, M., and Nelson, S., "Reliability of the Fugl-Meyer assessment of sensorimotor recovery following cerebrovascular accident," *Physical Therapy*, vol. 63, no. 10, p. 1606, 1983.

[39] Fasola, J. and Matarić, M., "Using socially assistive human-robot interaction to motivate physical exercise for older adults," in *To appear in Proceedings of the IEEE, Special Issue on Quality of Life Technology* (Kanade, T., ed.), 2012.

[40] Federico Posteraro, M., Mazzoleni, S., PhD, S., and MD, B., "Robot-mediated therapy for paretic upper limb of chronic patients following neurological injury," *REHABILITATION MEDICINE*, vol. 41, pp. 976–980, 2009.

[41] Fischler, M. and Bolles, R., "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM*, vol. 24, no. 6, pp. 381–395, 1981.

[42] Folio, M. and Fewell, R., "Peabody developmental motor scales and activity cards (manual): Hingham," *MA: Teaching Resources*, 1983.

[43] Folio, M. and Fewell, R., "Peabody developmental motor scales," *Austin, TX: Pro-Ed*, 2000.

[44] Fugl-Meyer, A., "Post-stroke hemiplegia assessment of physical properties.," *Scandinavian journal of rehabilitation medicine. Supplement*, vol. 7, p. 85, 1980.

[45] Fugl-Meyer, A., Jääskö, L., Leyman, I., Olsson, S., and Steglind, S., "The post-stroke hemiplegic patient. 1. a method for evaluation of physical performance.," *Scandinavian Journal of Rehabilitation Medicine*, vol. 7, no. 1, p. 13, 1975.

[46] Galloway, J., Ryu, J., and Agrawal, S., "Babies driving robots: self-generated mobility in very young infants," *Intelligent Service Robotics*, vol. 1, no. 2, pp. 123–134, 2008.

[47] Gockley, R. and MataríC, M., "Encouraging physical therapy compliance with a hands-off mobile robot," in *Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction*, pp. 150–155, ACM, 2006.

[48] Goncalves, L., DiBernardo, E., Ursella, E., and Perona, P., "Monocular tracking of the human arm in 3d," in *Proc. Int'l Conference Computer Vision*, pp. 764–770, August 1995.

[49] Hesse, S., Schulte-Tigges, G., Konrad, M., Bardeleben, A., and Werner, C., "Robot-assisted arm trainer for the passive and active practice of bilateral forearm and wrist movements in hemiparetic subjects," *Arch Phys Med Rehabil*, vol. 84, pp. 915–920, June 2003.

[50] Hurley, M., "Effectiveness of robot-assisted therapy on stroke patients with upper extremity impairment," *Physical Function*, p. 14, 2009.

[51] Institution, "Center for Disease Control and Prevention."

[52] Jackson, A., Culmer, P., Levesley, M., Makower, S., Cozens, J., and Bhakta, B., "Effector force requirements to enable robotic systems to provide assisted exercise in people with upper limb impairment after stroke," in *Rehabilitation Robotics (ICORR), 2011 IEEE International Conference on*, pp. 1–6, IEEE, 2011.

[53] Johnson, J., "Children, robotics, and education," *Artificial Life and Robotics*, vol. 7, no. 1, pp. 16–21, 2003.

[54] Jolliffe, I., *Principal Component Analysis, Series: Springer Series in Statistics, 2nd ed.* Springer, 2002.

[55] Kiguchi, K. and Fukuda, T., "A 3 DOF exoskeleton for upper limb motion assist: Consideration of the effect of bi-articular muscles," in *Robotics and Automation, 2004. Proceedings. ICRA'04. 2004 IEEE International Conference on*, vol. 3, pp. 2424–2429, IEEE, 2004.

[56] Koller-Hodac, A., Leonardo, D., Walpen, S., and Felder, D., "A novel robotic device for knee rehabilitation improved physical therapy through automated process," in *Biomedical Robotics and Biomechatronics (BioRob), 2010 3rd IEEE RAS and EMBS International Conference on*, pp. 820–824, IEEE, 2010.

[57] Konolige, K. and P., M., "Kinect calibration - technical - ros wiki." http://www.ros.org/wiki/kinect_calibration/technical, dec 2010.

[58] Kozima, H., Nakagawa, C., and Yasuda, Y., "Interactive robots for communication-care: A case study in autism therapy," in *ROMAN 2005*, pp. 341–346, August 2005.

[59] Krebs, H. and Hogan, N., "Therapeutic robotics: A technology push," *Proceedings of the IEEE*, vol. 94, no. 9, pp. 1727–1738, 2006.

[60] Krebs, H., Hogan, N., Aisen, M., and Volpe, B., "Robot-aided neurorehabilitation," *IEEE Trans. Rehab Eng.*, vol. 6, pp. 75–87, March 1998.

[61] Kwon, J., Kim, J., and Choi, H., "Classification of the EMG Signal using Cascaded Classifier," in *World Congress on Medical Physics and Biomedical Engineering 2006*, pp. 1222–1225, Springer, 2007.

[62] Lee, K., Jung, Y., Kim, J., and Kim, S., "Are physically embodied social agents better than disembodied social agents?: The effects of physical embodiment, tactile interaction, and people's loneliness in human-robot interaction," *International Journal of Human-Computer Studies*, vol. 64, no. 10, pp. 962–973, 2006.

[63] Loureiro, R., Amirabdollahian, F., Topping, M., Driessen, B., and Harwin, W., "Upper limb robot mediated stroke therapy gentle/s approach," *Autonomous Robots*, vol. 15, pp. 35–51, July 2003.

[64] Lum, P., Burgar, C., Shor, P., Majmundar, M., and der Loos, M. V., "Robot-assisted movement training compared with conventional therapy techniques for the rehabilitation of upper-limb motor function after stroke," *Arch Phys Med Rehabil*, vol. 83, pp. 952–959, July 2002.

[65] Lyle, R., "A performance test for assessment of upper limb function in physical rehabilitation treatment and research," *International Journal of Rehabilitation Research*, vol. 4, no. 4, p. 483, 1981.

[66] MacQueen, J., "Some Methods for Classification and Analysis of Multivariate Observations, 1967," in *Proceedings of 5 th Berkeley Symposium On Mathematical Statistics and Probability, Berkeley, University of California Press*, vol. 1, pp. 281–297, 1967.

[67] Mathiowetz, V., Volland, G., Kashman, N., and Weber, K., "Adult norms for the Box and Block Test of manual dexterity.," *The American journal of occupational therapy.: official publication of the American Occupational Therapy Association*, vol. 39, no. 6, p. 386, 1985.

[68] McDonnell, M., "Action Research Arm Test," *Australian Journal of Physiotherapy*, vol. 54, 2008.

[69] MHSportpflegemittel, "Mhsportpflegemittel - details - kniegoniometer." http://www.mhsportpflegemittel.de/de/details/kniegoniometer_/1386.aspx, accessed Feb 2012.

[70] Michaud, F., Salter, T., Duquette, A., Mercier, H., Lauria, M., Larouche, H., and Larose, F., "Assistive technologies and child-robot interaction," in *AAAI Spring Symposium on Multidisciplinary Collaboration for Socially Assistive Robotics*, 2007.

[71] MobilitySmart, "Welcome to mobility smart online." http://www.mobilitysmart.cc, Jan 2011.

[72] Mori, M., "The uncanny valley," *Energy*, vol. 7, no. 4, pp. 33–35, 1970.

[73] Morphopedics, "Shoulder impingement syndrome - morphopedic." http://morphopedics.wikidot.com/shoulder-impingement-syndrome, accessed Feb 2012.

[74] Murphy, M., Willén, C., and Sunnerhagen, K., "Kinematic variables quantifying upper-extremity performance after stroke during reaching and drinking from a glass," *Neurorehabilitation and neural repair*, vol. 25, no. 1, p. 71, 2011.

[75] Nanda, P., Smith, A., Gebregiorgis, A., and Brown, E., "Design and development of an upper extremity motion capture system for a rehabilitation robot," in *Engineering in Medicine and Biology Society, 2009. EMBC 2009. Annual International Conference of the IEEE*, pp. 7135–7138, IEEE, 2009.

[76] of Neurological Disorders, N. I. and Stroke, "Cerebral Palsy: Hope Through Research." NIH Publication Number 06-159, updated 7/13/07.

[77] OTA, Y., "Robot and mounting technology. a robot capable of autonomously walking by four feet. feasibility on a robot,'aibo'of the sony entertainment corporation.," *Electronic Packaging Technology*, vol. 15, no. 10, pp. 30–33, 1999.

[78] Park, S., Ji, S., Ryu, D., and Cho, H., "A smart and realistic chatting interface for gaming agents in 3-d virtual space," in *2008 International Conference on Games Research and Development 2008 (Cyber Games), To appear*, 2008.

[79] Parsons, T., *Voice and Speech Processing*. New York: McGraw-Hill, 1987.

[80] Patrizia, M., Claudio, M., Leonardo, G., and Alessandro, P., "A robotic toy for children with special needs: From requirements to design," in *Rehabilitation Robotics, 2009. ICORR 2009. IEEE International Conference on*, pp. 918–923, IEEE, 2009.

[81] Patton, J. L. and Mussa-Ivaldi, F. A., "Robot-assisted adaptive training: custom force fields for teaching movement patterns," *IEEE Trans Biomed Eng.*, vol. 21, pp. 636–646, April 2004.

[82] Platz, T., Pinkowski, C., van Wijck, F., Kim, I., di Bella, P., and Johnson, G., "Reliability and validity of arm function assessment with standardized guidelines for the Fugl-Meyer Test, Action Research Arm Test and Box and Block Test: a multicentre study," *Clinical rehabilitation*, vol. 19, no. 4, p. 404, 2005.

[83] Polana, R. and Nelson, R., "Low level recognition of human motion," in *Proc. IEEE Workshop Non-Rigid and Articulated Motion*, pp. 77–82, 1994.

[84] Powers, A., Kiesler, S., Fussell, S., and Torrey, C., "Comparing a computer agent with a humanoid robot," *Human Robot Interaction 2007*, pp. 145–152, 2007.

[85] Pransky, J., "Aibo-the no. 1 selling service robot," *Industrial robot: An international journal*, vol. 28, no. 1, pp. 24–26, 2001.

[86] Press, W., *Numerical recipes in FORTRAN: the art of scientific computing*, vol. 1. Cambridge Univ Pr, 1992.

[87] Rabiner, L., "A tutorial on hidden Markov models and selected applications in speech recognition," *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257–286, 1989.

[88] Rasmussen, C. and Williams, C., "Gaussian processes for machine learning," 2006.

[89] Rehg, J. and Kanade, T., "Model-based tracking of self-occluding articulated objects," in *Proc. Int'l Conference Computer Vision*, pp. 612–617, 1995.

[90] Reinkensmeyer, D. J., Schmit, B. D., and Rymer, W. Z., "Assessment of active and passive restraint during guided reaching after chronic brain injury," *Ann Biomed Eng.*, vol. 27, pp. 805–814, Nov-Dec 1999.

[91] Riener, R., Guidali, M., Keller, U., Duschau-Wicke, A., Klamroth, V., and Nef, T., "Transferring armin to the clinics and industry," *Topics in Spinal Cord Injury Rehabilitation*, vol. 17, no. 1, pp. 54–59, 2011.

[92] Robins, B., Dautenhahn, K., Te Boekhorst, R., and Billard, A., "Effects of repeated exposure to a humanoid robot on children with autism," *Designing a More Inclusive World*, pp. 225–236, 2004.

[93] Robotis, "Robotis - [darwin-op] open platform humanoid project." http://www.robotis.com/xe/37937, dec 2010.

[94] Rothe, I., Susse, H., and Voss, K., "The method of normalization to determine invariants," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 18, no. 4, pp. 366–376, 1996.

[95] Russell, D., Rosenbaum, P., Cadman, D., Gowland, C., Hardy, S., and Jarvis, S., "The gross motor function measure: a means to evaluate the effects of physical therapy," *Developmental Medicine & Child Neurology*, vol. 31, no. 3, pp. 341–352, 1989.

[96] Sanford, J., Moreland, J., Swanson, L., Stratford, P., and Gowland, C., "Reliability of the Fugl-Meyer assessment for testing motor performance in patients following stroke," *Physical Therapy*, vol. 73, no. 7, p. 447, 1993.

[97] Scassellati, B., "How social robots will help us to diagnose, treat, and understand autism," *Robotics Research*, pp. 552–563, 2007.

[98] Shapiro, L. G. and Stockman, G. C., *Computer Vision*. Prentice Hall, 2001. retrieved from Wikipedia.

[99] Shen, D. and Ip, H., "Generalized affine invariant image normalization," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 19, no. 5, pp. 431–440, 1997.

[100] Shen, D., Ip, H., Cheung, K., and Teoh, E., "Symmetry detection by generalized complex (gc) moments: a close-form solution," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 21, no. 5, pp. 466–476, 1999.

[101] Solak, E., Murray-Smith, R., Leithead, W., Leith, D., and Rasmussen, C., "Derivative observations in gaussian process models of dynamic systems," 2003.

[102] Stone, E. and Skubic, M., "Evaluation of an inexpensive depth camera for passive in-home fall risk assessment," in *Pervasive Computing Technologies for Healthcare (PervasiveHealth), 2011 5th International Conference on*, pp. 71–77, IEEE, 2011.

[103] Suzuki, S. and others, "Topological structural analysis of digitized binary images by border following," *Computer Vision, Graphics, and Image Processing*, vol. 30, no. 1, pp. 32–46, 1985.

[104] Takeuchi, J., Kushida, K., Nishimura, Y., Dohi, H., Ishizuka, M., Nakano, M., and Tsujino, H., "Comparison of a humanoid robot and an on-screen agent as presenters to audiences," in *Intelligent Robots and Systems, 2006 IEEE/RSJ International Conference on*, pp. 3964–3969, IEEE, 2006.

[105] Tapus, A., Mataric, M., and Scassellati, B., "Assistive robotics," *IEEE Robotics & Automation Magazine*, p. 36, 2007.

[106] Trevor, A., Park, H., Howard, A., and Kemp, C., "Playing with toys: towards autonomous robot manipulation for therapeutic play," in *Proceedings of the 2009 IEEE international conference on Robotics and Automation*, pp. 2586–2592, IEEE Press, 2009.

[107] Tuzun, H., Yilmaz-Soylu, M., Karakus, T., Inal, Y., and Kizilkaya, G., "The effects of computer games on primary school students' achievement and motivation in geography learning," *Computers & Education*, vol. 52, no. 1, pp. 68–77, 2009.

[108] van der Lee, J., de Groot, V., Beckerman, H., Wagenaar, R., Lankhorst, G., and Bouter, L., "The intra-and interrater reliability of the

action research arm test: A practical test of upper extremity function in patients with stroke* 1," *Archives of physical medicine and rehabilitation*, vol. 82, no. 1, pp. 14–19, 2001.

[109] Volpe, B. T., Krebs, H. I., Hogan, N., Edelstein, L., Diels, C., and Aisen, M., "A novel approach to stroke rehabilitation: robot-aided sensorimotor stimulation," *Neurology*, vol. 54, pp. 1938–1944, May 2000.

[110] Von Mises, R., "Mathematical theory of probability and statistics," *Mathematical Theory of Probability and Statistics, New York: Academic Press, 1964*, vol. 1, 1964.

[111] Wakiji, E., "Mapping the literature of physical therapy.," *Bulletin of the Medical Library Association*, vol. 85, no. 3, p. 284, 1997.

[112] Wang, H., Liao, H., and Hsieh, C., "Reliability, sensitivity to change, and responsiveness of the peabody developmental motor scales–second edition for children with cerebral palsy," *Physical Therapy*, vol. 86, no. 10, p. 1351, 2006.

[113] White, D., Catanzaro, M., and Kraft, G., "An approach to the psychological aspects of multiple sclerosis: A coping guide for healthcare providers and families," *Neurorehabilitation and Neural Repair*, vol. 7, no. 2, p. 43, 1993.

[114] WHO, "International Classification of Functioning, Disability and Health (ICF)," 2001.

[115] Williams, S., "Visual arctic navigation: techniques for autonomous agents in glacial environments," 2011.

[116] Wilson, A. and Benko, H., "Combining multiple depth cameras and projectors for interactions on, above and between surfaces," in *Proceedings of the 23nd annual ACM symposium on User interface software and technology*, pp. 273–282, ACM, 2010.

[117] Wolf, S., Thompson, P., Morris, D., Rose, D., Winstein, C., Taub, E., Giuliani, C., and Pearson, S., "The excite trial: attributes of the wolf motor function test in patients with subacute stroke," *Neurorehabilitation and neural repair*, vol. 19, no. 3, pp. 194–205, 2005.

[118] Yule, G., "An introduction of the theory of statistics. Charles Griffin & Co," *London UK Xi Xj abcd*, vol. 1, no. 1, p. 1, 1911.